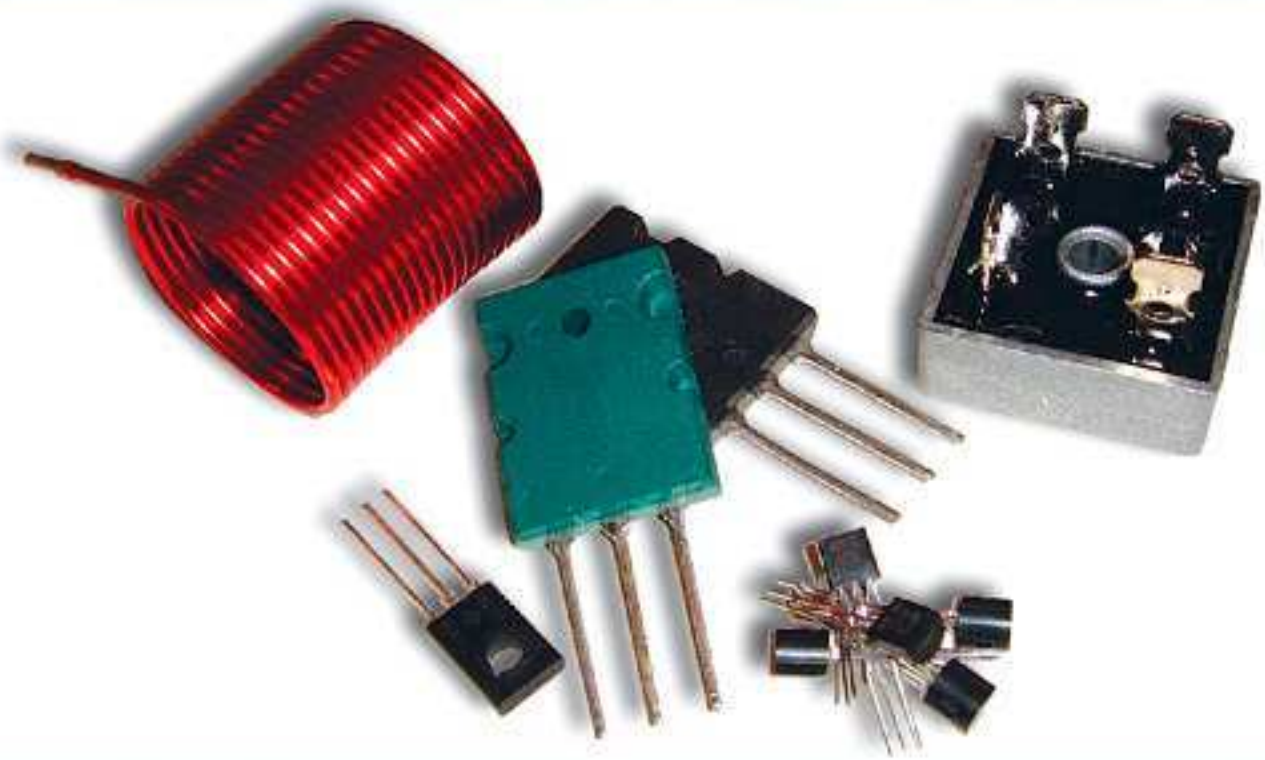
SECOND EDITION

# Self on Audio



Douglas Self

Newnes

VISIT…

Self on Audio

This Page is Intentionally Left Blank

# Self on Audio

Second Edition

Douglas Self

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Newnes is an imprint of Elsevier

Newnes is an imprint of Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP
30 Corporate Drive, Burlington, MA 01803

First published 2006

Copyright © 2006, Douglas Self. All rights reserved

The right of Douglas Self to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988

No part of this publication may be reproduced in any material form (including photocopying or storing in any medium by electronic means and whether or not transiently or incidentally to some other use of this publication) without the written permission of the copyright holder except in accordance with the provisions of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London, England W1T 4LP. Applications for the copyright holder's written permission to reproduce any part of this publication should be addressed to the publisher

Permissions may be sought directly from Elsevier's Science and Technology Rights Department in Oxford, UK: phone: (+44) (0) 1865 843830; fax: (+44) (0) 1865 853333; e-mail: permissions @ elsevier.co.uk. You may also complete your request on-line via the Elsevier homepage (http://www.elsevier.com), by selecting 'Customer Support' and then 'Obtaining Permissions'

**British Library Cataloguing in Publication Data**
A catalogue record for this book is available from the British Library

**Library of Congress Control Number:** 2006927676

ISBN–13: 978-0-7506-8166-7
ISBN–10: 0-7506-8166-7

For information on all Newnes publications visit our web site at www.newnespress.com

Printed and bound in *Great Britain*

06 07 08 09 10 10 9 8 7 6 5 4 3 2 1

Working together to grow libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER BOOK AID International Sabre Foundation

# Contents


Introduction viii

**Preamplifiers and related matters**

1 Advanced preamplifier design (November 1976) 1

2 High-performance preamplifier (February 1979) 13

3 Precision preamplifier (October 1983) 24

4 Design of moving-coil head amplifiers (December 1987) 37

5 Precision preamplifier '96, Part I (July/August 1996) 49

6 Precision preamplifier '96, Part II (September 1996) 67

7 Overload matters (February 1997) 86

8 A balanced view, Part I (April 1997) 90

9 A balanced view, Part II (May 1997) 103

10 High-quality compressor/limiter (December 1975) 117

11 Inside mixers (April 1991) 127

12 Electronic analogue switching, Part I: CMOS gates (January 2004) 142

13 Electronic analogue switching, Part II: discrete FETs (February 2002) 155

**Power Amplifiers**

14 Sound mosfet design
(September 1990) 170

15 FETs versus BJTs: the linearity competition
(May 1995) 181

16 Distortion in power amplifiers, Part I: the sources of distortion
(August 1993) 187

17 Distortion in power amplifiers, Part II: the input stage
(September 1993) 201

18 Distortion in power amplifiers, Part III: the voltage-amplifier stage
(October 1993) 217

19 Distortion in power amplifiers, Part IV: the power amplifier stages
(November 1993) 232

20 Distortion in power amplifiers, Part V: output stages
(December 1993) 251

21 Distortion in power amplifiers, Part VI: the remaining distortions
(January 1994) 265

22 Distortion in power amplifiers, Part VII: frequency compensation and real designs
(February 1994) 279

23 Distortion in power amplifiers, Part VIII: Class A amplifiers
(March 1994) 294

24 Power amplifier input currents and their troubles
(May 2003) 310

25 Diagnosing distortions
(January 1998) 318

26 Trimodal audio power, Part I
(June 1995) 333

27 Trimodal audio power, Part II
(July 1995) 347

28 Load-invariant audio power
(January 1997) 363

29 Common-emitter power amplifiers: a different perception? (July 1994) 380

30 Few compliments for non-complements (September 1995) 391

31 Loudspeaker undercurrents (February 1998) 400

32 Class distinction (March 1999) 404

33 Muting relays (July 1999) 421

34 Cool audio power (August 1999) 434

35 Audio power analysis (December 1999) 448

Index 463

# Introduction

This book is a collection of the articles I wrote for the journal *Wireless World* (now *Electronics World*) between the years 1979 and 1999. The vast majority of these deal with hi-fi preamplifiers and power amplifiers, and this book concentrates on this field.

In the last twenty-five years the scope of technology available to the audio designer has greatly widened. At the beginning of this period the only choice in preamp design was between discrete transistor stages and the relatively new op-amps. The latter had dubious characteristics as regards noise and distortion—particularly crossover distortion. This most unloved of audio defects was tolerated in power amplifiers because it had to be, but there was considerable resistance to incorporating it in preamplifiers. At this time no-one would have considered using valve circuitry in a new design.

My association with this influential and much-loved magazine actually began with a design for a compressor/limiter in 1975, which started life as my third-year project at university. This design is not reproduced here as it is unarguable that it has been overtaken by advances in technology.

This debut was followed by what I called 'An Advanced Preamplifier' in 1976. The Advanced Preamplifier certainly gave (and gives—I still have the prototype) exemplary performance, obtained by making each stage a discrete-component operational amplifier. This made necessary the use of dual IC regulators to produce $+/-15$ V, and at the time the cost of this power-supply scheme was significant.

As a reaction to this complexity, I decided to try my hand at what might be called 'traditional' discrete circuitry in a preamplifier, and this became the 'High Performance Preamplifier' published in 1979, though actually designed nearly two years earlier. It was conceived in an era when op-amps were still regarded with considerable suspicion by designers seeking the best possible audio performance. In the search for simplicity a single supply rail was used, without regulation, but with a simple RC filter after the reservoir capacitor to reduce ripple to a manageable 50 mV or so. Experiment had proved that this minimal-cost arrangement could give hum and noise results that were as good as those yielded by the dual-IC-regulator approach.

In contrast, the Precision Preamplifier of 1983 was designed at a time when the remarkable 5534/5532 op-amps had become available at reasonable prices. Since they delivered very low noise with almost unmeasurable

distortion, it was clearly time to try a 'third way' as regards preamp design. Having explored discrete op-amps, and conventional discrete circuitry, an IC solution was an obvious next step. The return to op-amps meant a return to dual power supplies, but this was a small price to pay for the convenience of dual rails. This design later gained a moving-coil head amplifier. I had designed several of these stages before, at least two of which made their way into commercial production, but this was the first version that got both noise and distortion down to what I considered to be acceptably low levels. The salient features are the discrete transistor input devices which then, and indeed now, provide the best possible noise figure. At the time many head amps were outboard units, often relying on battery power, presumably to sidestep intractable ground-loop problems. However, no difficulties were found in grafting this design onto existing preamplifiers.

Some years later, having devoted much time in between to power amplifier design, I felt the call to take another look at preamplifiers. My last design was twelve years old, and it seemed likely that some significant improvements could be made. Much thought and a lot of calculation and simulation led to the 'Precision Preamp 96' articles, including in-depth mathematical modelling of the noise generated by the RIAA stage. This allowed each noise contribution to be studied independently, and permitted comparison between the actual noise and the theoretical minimum. The latter is rather sensitive to the exact assumptions made. It was also possible to discover why op-amps that appeared to be quieter than the 5534 in theory, were actually slightly noisier in practice. The answer was that op-amp bias-cancellation networks maybe great for DC precision, but the extra common-mode noise they generate in audio circuitry is just an embarrassment. The 96 preamp is in fact not so much an updated version as a thorough re-design, with only the moving-coil input amp remaining essentially unaltered. It demonstrated, amongst other things, that obtaining an interchannel separation of 100 dB on a stereo PCB is perfectly possible with careful component and track layout.

In 1990 I had again turned my attention to power amplifiers. For many years I had felt that the output stages of power amplifiers presented very great possibilities for creative design, and so I explored some of them. One of the first difficulties I met with was the problem of determining how much of the overall distortion was produced in the small-signal sections, and how much was generated by the output stage. Traditionally the latter was regarded as the major source of distortion, but there was very little published research to back this up, and so I attacked the problem myself. When I began it was not clear if there were two, twenty, or two hundred significant distortion mechanisms, but after a good deal of study it suddenly became clear that seven or eight were sufficient to explain all the observable distortion. This is not to say that there are not other distortion

mechanisms—there almost certainly are—but the non-linearities they produce are currently below the level of practical measurement with THD analysers. When output stage distortion has been banished forever (and it has to be said there is no sign of this happening in the immediate future) then it may be time to dig into the deeper levels of non-linearity. It quickly became clear that by taking a few simple circuit precautions, it was possible to design amplifiers with very much lower distortion than the norm. Such amplifiers, with their very low THD figures, are rather distinct from average designs, and so I looked around for a suitable name. Once the critical factors are identified, designing an low-distortion amplifier becomes more a matter of avoiding mistakes rather than being brilliant, so I decided to call them 'blameless' amplifiers, rather than 'hyper-linear' or something similar, to emphasise this. The results and conclusions of this major investigation were published in eight parts as 'Distortion in Power Amplifiers'. 'Distortion Residuals' followed this up, providing a visual guide to the appearance of the various distortion mechanisms on the oscilloscope screen.

These endeavours built up to a substantial body of information on just how to minimise amplifier distortion, and two designs that exemplify this are included in the 'Distortion In Power Amplifiers' articles, one working in Class-B and the other in Class-A. This foundation of knowledge simply begged to be put to further use, and so two major power amplifier projects were created; the trimodal amplifier based on Class-A, and the load-invariant amplifier in relatively conventional Class-B.

The trimodal amplifier demonstrated how to make a Class-A amplifier that coped gracefully with varying load impedances. This project had its roots in an insistent demand for a PCB for the Class-A power amplifier presented in the last part of 'Distortion in Power Amplifiers'. I find it goes against the grain to reproduce a design without trying to improve it, and the trimodal article was the result. My first intention was to demonstrate how a Class-A amplifier could, with appropriate design, move gracefully into a relatively linear version of Class-AB when the load impedance became too low for Class-A operation to be maintained, rather than clipping horribly as some configurations do. Improvements were also made in the noise and DC offset performance of the basic amplifier. I was concerned to guard against catastrophic currents flowing if there were errors in building the quiescent-current controller, so a safety network was added to set an upper limit on the bias voltage. It was simple to make the amplifier switchable between A and B by changing the limiting value of this second bias circuit, and the trimodal was born.

The load-invariant power amplifier project was a direct development of the work done on amplifier distortion. Power amplifiers always give worse distortion into lower load impedances. For bipolar output devices, as the load value drops from 100 Ω to about 8 Ω, the crossover distortion

increases steadily and predictably. However, at about 8 Ω (depending on transistor characteristics) an extra low-order distortion appears that can easily double the THD at 4 Ω. I decided to see to what extent I could thwart this extra distortion, aiming to produce the first semiconductor amplifier that gave exactly the same THD at 4 Ω as it did at 8 Ω. While it did not prove possible to quite attain this, I did get reasonably close. This design seems to have generated a lot of interest.

A few of my articles have been written in reaction to contributions to *Electronics/Wireless World* that suggested promising new approaches, or that I simply found intriguing. Investigating a particular amplifier topology takes a lot of time and effort, but preparing the results for publication does not add a great deal to this, so some more articles resulted. They may not have advanced the art of audio greatly, but they did explore a few paths which would otherwise have remained untrod. Two examples are given in this book: 'Common-Emitter Amplifiers' and 'Two-Stage Amplifiers'. In neither case were the results sufficiently encouraging for me to proceed further with the concepts involved. When the idea that loudspeakers could, in certain circumstances, draw much more current from an amplifier than its impedance curve suggested first came to my notice in the mid-1980s, I must confess I felt a degree of scepticism. I was wrong; the effect is real, though its relevance to real-life signals rather than artificial stimulus waveforms is rather doubtful. The abnormally high currents that flow are provoked by using the stored energy in the circuit elements, such as the inertia of the speaker cone. This requires a rectangular stimulus waveform with rapid full-amplitude transitions, carefully timed to catch the speaker resonance at its worst moment, and the difficulty is that real waveforms do not have these. Eventually I got around to putting the idea to the test, using an electrical analogue of a speaker system. The article 'Excess Speaker Currents' describes the effect and how to produce it.

When semiconductors were first applied to audio amplification, the choice of operating mode was simple: Class-A or Class-B. It took several years before proper complementary pairs of output devices were available, but it was clear that they allowed a good deal more flexibility in the design of output stages, and variations on the standard configurations began to appear, becoming more radical as time went on and the technology of audio developed. The 1960s gave us Class-D, though the rudimentary versions available then were not much of a gift. The 1970s saw the advent of the Blomley concept, current-dumping, and, significantly, Class-G. In this situation it was inevitable that extra letters of the alphabet would be called in to describe new methods of operation, though it was clear that calling something, say, 'Class-Y' said nothing about how it operated, and the prospect of trying to remember what an alphabet soup of 26 (or more) class letters actually represented was not enticing. With this in mind, I produced the article 'class distinction' which attempts to simplify amplifier

classification into simple combinations of A, B, C and D in such a way that at least some information about the mode of action is given. I will not pretend that I expect my classification system to sweep the world overnight. Should it fail to sweep the world at all, I think the article is still useful because it allows the generation of a matrix of amplifier types, some of which no-one has got around to inventing yet. Give me time.

Power FETs first began to reach the market in the mid-1970s, and as so often with new technology, they were claimed to be superior to existing methods in just about every possible way. However, experience soon showed that FETs were not dramatically more linear than bipolar transistors, nor were they inherently short-circuit proof. There appeared, however, to be definite advantages in their high bandwidth and freedom from carrier-storage effects. Initially I was intrigued by the possibility that power FETs, with their much-advertised speed and bandwidth, would allow the implementation of various ingenious output stages involving local feedback that in bipolar format had proved difficult or impossible to stabilise, apparently due to the slowness of the output devices. The results were not encouraging: any increase in stability due to the faster devices was more than outweighed by their tendency to parasitic oscillation when used in anything other than the simplest of configurations.

Most serious power amplifiers are fitted with a muting relay that disconnects the electronics from the loudspeaker at turn-on and turn-off, to reduce thumps and bangs. The same relay is used to protect the loudspeaker from incineration if the amplifier suffers a fault which puts a large DC voltage on the output. Because of the safety implications, this relay must operate promptly and reliably, and designing its control circuitry is not trivial. My article on relay control delved deeply into the control circuit design, with special emphasis on a rapid relay response to dangerous conditions or intrusive transients; this is not, as it might appear, merely a matter for the relay designer. Apparently tiny details of circuit design have a major effect.

The final two articles in this book introduce a new way of displaying amplifier efficiency and power dissipation that I call 'power partition diagrams'. The first one dealt with many different kinds of amplifier, and studied how they disposed of the power involved in driving resistive and reactive loads with sine waves. This produced the interesting conclusion that a proper appreciation of peak transistor dissipations into real loads with phase-shift could be the most crucial factor in determining amplifier reliability. Whenever sine waves are use for testing—and there are many good reasons why they should be used sometimes—the criticism is likely to be levelled that this is unrealistic, which of course it is. The second article therefore looked at the statistics of music (which are surprisingly obscure) and ways of calculating true power dissipations from the results. This showed, amongst other things, that Class-A amplifiers are in reality

not more than 1% efficient. This raises serious questions about their desirability in an energy-conscious world.

In the course of the investigations that led to these articles, I found over and over again that the conventional wisdom on power amplifiers was more conventional than wise. Some examples are given here, though they may not make much sense until you have read the relevant article. Some statements turned out to be half-true: an example is the widespread assumption that a current-source loaded Voltage Amplifier Stage (VAS) gives current drive to the output devices. The reality is much more complex; the impedance might be high at low frequencies, but the Miller capacitor around the VAS causes the impedance to fall with frequency until it is a few kilo ohms at the top of the audio band. This hardly counts as a current source. The drive point is also loaded by the non-linear input impedance of the output stage, which complicates matters further. There is much more to most questions about amplifiers than at first meets the eye.

Some long-held beliefs turned out to be completely wrong, though plausible in theory and workable in practice. The best illustration of this is the universal belief that the crucial parameter in biasing a Class-B output stage to minimise crossover distortion is the quiescent current. In actual fact the critical factor is the voltage across the emitter resistors. If the value of these are changed, then the quiescent current can be radically altered although the amplifier remains at the same optimal bias point, because the voltage drop is unchanged. However, emitter resistor values are rarely changed in this way, so setting up for a given current is the same as setting up for a given voltage. The difference is unimportant if you are simply repairing or adjusting amplifiers, but vital if you seek to understand how they work.

In some cases there was no argument about the distortion mechanism operating, but very little, if any, published information quantifying the size of the effect. This applied to most of the amplifier distortions examined, and it took a little thought to develop ways of measuring each one separately, wondering in each case why the relatively simple test had apparently never been done before. It may be, of course, that parts of this work have also been done by various audio manufacturers, who have every reason for keeping their private research to themselves.

This collection of articles is not totally exhaustive, as various topics that have already been fully expounded in *The Audio Power Amplifier Design Handbook* have been omitted. There is some overlap in the material relating to distortion, but that is unavoidable if each book is to stand alone. The articles here are not in chronological order as it is more useful to keep preamplifiers and power amplifier material grouped together. Writing the articles reproduced here has been a stimulating intellectual journey.

I hope that reading them will share some of the sense of discovery that I felt, and that this collection will be both useful and entertaining to all those concerning themselves with audio electronics.

This Page is Intentionally Left Blank

# 1 Advanced preamplifier design

## A no-compromise circuit with noise gating

*November 1976*

This was my first preamplifier design, conceived in the first year that I worked in the audio industry. The two gain-control solution to the dynamic-range problem was inspired by industry practice at the time – notably Radford and Cambridge Audio. I thought the rumble-gate concept was rather clever, but I seem to have been in a minority of one, as it was never even mentioned in the correspondence that followed. The level monitoring circuitry was perhaps a bit over the top, and I have to admit that I never once saw the 'CLIP' LEDs come on in real use. The clip-detect circuit proved extremely useful in other applications however, and I have been using it ever since. There is also a bit of an infelicity in the '1V Peak' driver circuit, where the full output swing of one of the discrete op-amps is potentially applied to the trigger input of a 555 timer. Still, it never failed.

You may be wondering why the coupling capacitors are in most cases rather small. This is because electrolytics were not quite as reliable then as they are now, and so I wanted to use as few as possible.

With the glorious clarity of hindsight, the LED biasing of current-sources was quite unnecessary, as the discrete op-amps are not that sensitive to their internal operating conditions. Nevertheless, it made for a very pretty PCB.

This preamplifier design offers a distortion figure of below 0.002%, an overload margin of around 47 dB, and a signal-to-noise ratio of about 71 dB for the disc amplifier. A novel noise gate mutes the output when no signal is presented to the disc input and conversely, by using the subsonic information present on record pressings, eliminates the problem of muting low-level signals.

This article describes a stereo pre-amplifier that equals or exceeds the performance of many of those available. The circuit incorporates a novel method of muting the signal path, when the disc input is quiescent, by using a noise gate that never mutes a wanted low-level signal.

Many of the important performance factors, such as signal-to-noise ratio, overload margin, and accuracy of the RIAA equalization, are essentially defined by the design of the disc input circuitry. This therefore merits close attention. The best attainable s/n ratio for a magnetic cartridge feeding a bipolar transistor stage with series feedback is about 71 dB with respect to a 2 mV r.m.s. input at 1 kHz, after RIAA equalization. This has been clearly demonstrated by Walker.[1] The equivalent amplifier stage with shunt feedback gives an inferior noise performance over most of the audio band due to the rise in cartridge source impedance with frequency. This limits the maximum s/n ratio after equalization to about 58 dB. These facts represent a limit to what the most advanced disc input stage can achieve.

Overload margin appears to be receiving little attention. The maximum velocities recorded on disc seem to be steadily increasing and this, coupled with improved cartridges, means that very high peak voltages are reaching disc inputs. Several writers have shown that short-term voltages of around 60–80 mV r.m.s. are possible from modern disc and cartridges, and higher values are to be expected.[2,3] This implies that to cater for signal maxima, a minimum overload margin of 32 dB with respect to 2 mV r.m.s. at 1 kHz is essential. Obviously a safety factor on top of this is desirable. However, most pre-amplifiers at the top end of the market provide around 35–40 dB only. There are certain honourable exceptions such as the Technics SU9600 control amplifier which achieves an overload margin of 54 dB, mainly by the use of a staggeringly high supply of 136 V in the disc input amplifier. The Cambridge P50/110 series offers a margin in excess of 60 dB by the artifice of providing unity-gain buffering, for correct cartridge loading, but no amplification before the main gain control. This allows the use of an 18 V supply rail, but does limit the maximum s/n ratio.

The overload margin of a pre-amplifier is determined by the supply voltage which sets the maximum voltage swing available, and by the amount of amplification that can be backed-off to prevent overload of subsequent stages. Most preamplifiers use a relatively high-gain disc input amplifier that raises the signal from cartridge level to the nominal operating level in one jump. Low supply voltages are normally used which reduce static dissipation and allow the use of inexpensive semiconductors. The gain

control is usually placed late in the signal path to ensure low-noise output at low volume settings. Given these constraints, the overload performance is bound to be mediocre, and in medium-priced equipment the margin rarely exceeds 30 dB. If these constraints are rejected, the overload margin of the system can be improved.

Two separate gain controls remove the most difficult compromise, which is the placement of the volume control. This approach is exemplified in the Radford ZD22 and the Cambridge P60 circuitry. One gain control is placed early in the signal path, preceded by a modest amount of gain. Cartridges of high output can be accommodated by the use of this first control. The second is placed late in the pre-amplifier and is used as a conventional volume control, see Figure 1.

The other performance criterion which is largely defined by the disc input circuitry is frequency response, as defined by the accuracy of the RIAA equalization. Assuming that the relevant amplifying stage has sufficient open-loop gain to cope with the bass boost required, the accuracy of the equalization depends entirely on the time constants within the feedback loop. Careful design, and the use of close-tolerance components can assure an accurate response to within $\pm 0.2$ dB from 30 Hz to 20 kHz.

Pre-amplifier distortion seems to have received little attention compared with that generated by power amplifiers, perhaps because the former has traditionally been much lower. However, power amplifiers, with such low THD that the residual harmonics can no longer be extracted from the noise at normal listening levels, are now commonplace, particularly with the advent of techniques such as current dumping. This desirable state of affairs unfortunately does not extend to pre-amps, which in general produce detectable distortion at nominal operating levels, usually between 0.02% and 0.2%. In this design the THD at 1 kHz is less than 0.002% even at 25 dB above the nominal operating level of 0 dBm. A Sound Technology 1700 A distortion measurement system was used during development.

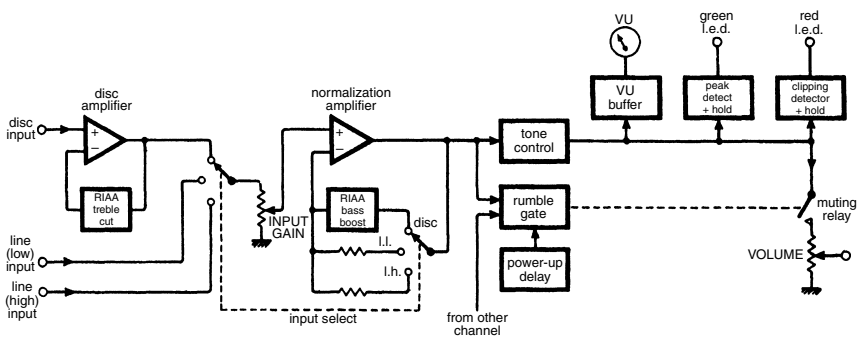



**Figure 1** *Block diagram of the complete circuit. Two gain controls are used in the signal path to allow a substantial increase in overload margin.*

At this point it is convenient to consider the noise gate principle. When the pre-amplifier is being used for disc reproduction the output from each channel is continuously sampled to determine if a signal is present; if nothing is detected within a specified time interval, dependent on the previous signal levels received, the pre-amplifier is muted by the opening of a reed relay in series with the output signal path. This allows only power amplifier noise to reach the loudspeakers and considerably reduces the perceived noise generated by a quiescent sound system. Noise in the quiescent state is particularly noticeable when headphones are in use. The reed relay is also used to prevent switch-on transients from reaching an external power amplifier. So far this circuit appears to be a fairly conventional noise gate. The crucial difference is that signals from disc that have not been subjected to rumble filtering are always accompanied by very low frequency signals generated by record ripples and small-scale warps. Even disc pressings of the highest quality produce this subsonic information, at a surprisingly high level, partly due to the RIAA bass boosting. The l.f. component is often less than 20 dB below the total programme level but this is quite sufficient to keep the pre-amplifier unmuted for the duration of a l.p. side. The preamplifier is unmuted as soon as the stylus touches the disc, and muted about a second after it has been raised from the run-out groove. This delay can be made short because the relative quiet at the start of the run-out groove is sensed and stored. The rumble performance of the record deck is largely irrelevant because virtually all of the subsonic information is generated by disc irregularities.

## Audio circuitry

A detailed block diagram of the pre-amplifier is shown in Figure 1, and Figure 2 shows the main signal path. The disc input amplifier uses a configuration made popular by Walker, but the collector load of the second transistor is bootstrapped. This increases the open-loop gain and hence improves the closed-loop distortion performance by a factor of about three to produce less than 0.002% at an output of 6.5 V r.m.s. (1 kHz). This stage gives a s/n ratio (ref 2 mV) of about 70 dB and a gain of 15 at 1 kHz. This is sufficient to ensure that the noise performance is not degraded by subsequent stages of amplification. The maximum output of this stage before clipping is about 6.5 V r.m.s. and the nominal output is 30 mV r.m.s. Because this is the only stage before the input gain control, these two figures set the overload margin at 47 dB. To ensure that this overload margin is maintained at high frequencies, the treble-cut RIAA time-constant is incorporated in the feedback loop. This leads to slightly insufficient cut at frequencies above 10 kHz because the gain of the stage cannot fall below unity, and hence fails to maintain the required 6 dB/octave fall at

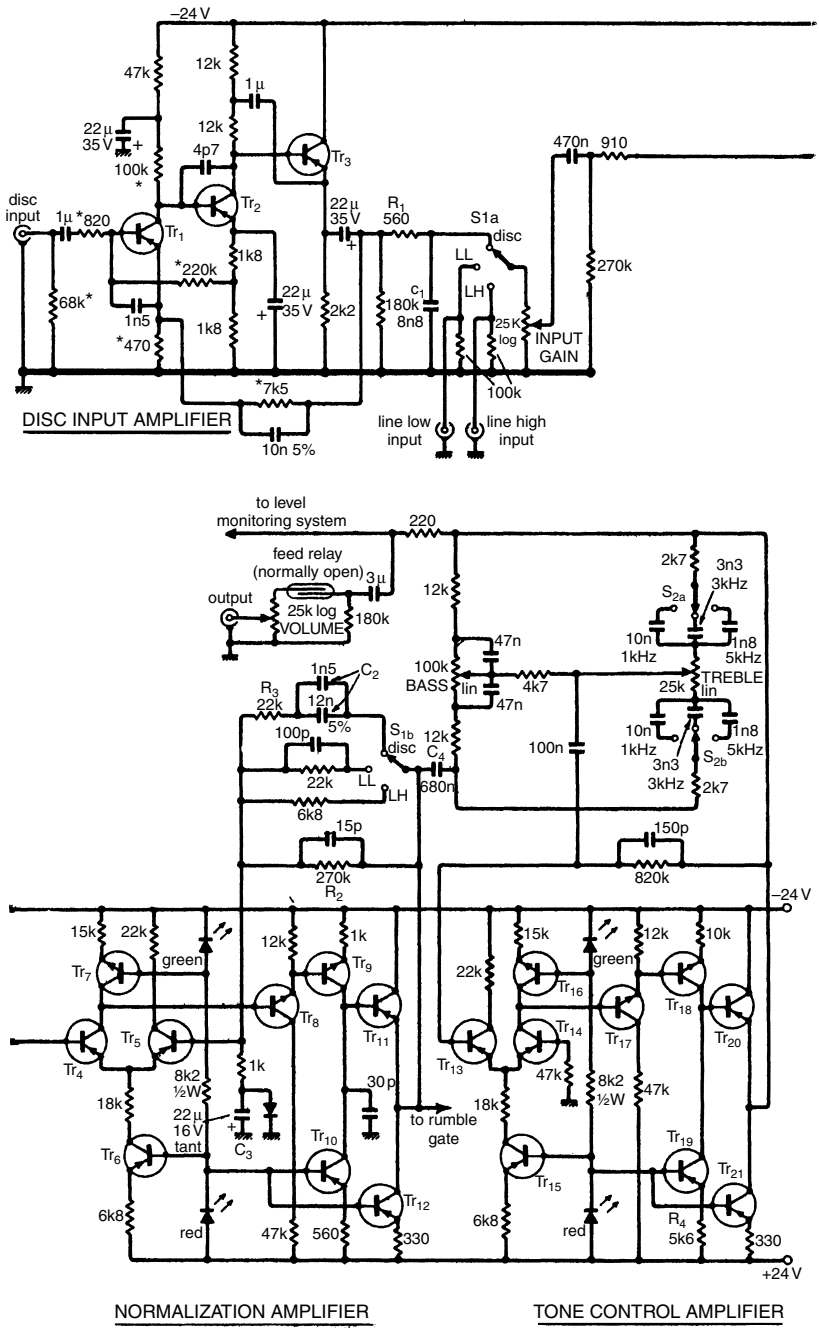



**Figure 2** *Circuit diagram of the signal path. Constant-current sources are biased from a l.e.d./resistor chain for improved thermal stability.*

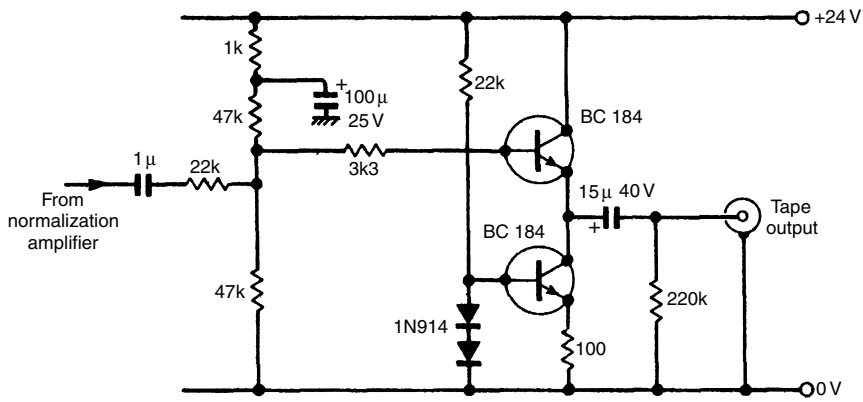



**Figure 3** *Tape output circuit. The smallest allowable load impedance for an undistorted output is about* 2.2 k$\Omega$. *Line inputs of the pre-amplifier are suitable for playback purposes.*

the top of the audio spectrum. This is exactly compensated for outside the feedback loop by the low-pass filter $R_1$ $C_1$, which also helps to reject high frequencies above the audio band.

For convenience I have referred to the next stage of the circuit as the normalization amplifier because signals leaving this should be at the nominal operating level of 0 dBm by manipulation of the input gain controls. Separate controls are provided for each channel to allow stereo balance. A later ganged control is used for volume setting and causes no operational inconvenience. In the disc replay mode, the normalization amplifier provides the RIAA bass boost, by the feedback components $R_{2+3}$ and $C_2$. Two line inputs are also provided; line low requiring 30 mV and line high 100 mV to give 0 dBm from the normalization stage with the input gain control fully advanced. When these inputs are selected, the feedback networks are altered to adjust the gain and give a flat frequency response. Ultrasonic filters are incorporated to ensure stability and aid r.f. rejection. Capacitor $C_3$ in the feedback arm reduces the gain to unity at d.c. for good d.c. stability. If a fault causes the amplifier output to saturate positively the capacitor is protected by a diode which has no effect on the distortion performance.

The circuitry of the normalization amplifier is complicated because its performance is required to be extremely high. The harmonic distortion is far below 0.002% at the maximum output of 14.5 V r.m.s. which is 25 dB above nominal operating level. This large amount of preamplifier headroom allows gross preamplifier overload before clipping. The input stage of the amplifier is a differential pair with a constant-current source for good common-mode rejection. The operating currents are optimized for

good noise performance, and the output is buffered by an emitter-follower. The main voltage amplifier, $Tr_9$ has a constant-current collector load so that high voltage gain at low distortion can be obtained. This performance is only possible if the stage has very little loading so it is buffered by the active-load emitter-follower. The various current sources are biased by a l.e.d.-resistor chain because the forward voltage drop of an l.e.d. has a negative temperature coefficient that approximates closely to that of a silicon transistor $V_{be}$ drop. Hence, this method provides exceptionally stable d.c. conditions over a very wide temperature range.

After the normalization stage the signal is applied to a tone-control circuit based on the Baxandall network. The main limitation of the Baxandall system is that the turnover frequency of the treble control is fixed. In contrast, the bass control has a turnover frequency that decreases as the control nears the flat position. This allows a small amount of boost at the low end of the audio spectrum to correct for transducer shortcomings. The equivalent adjustment at the high end of the treble spectrum is not possible because boost occurs fairly uniformly above the turnover frequency for treble control settings close to flat. In this circuit the treble turnover frequency has been given three switched values which have proved useful in practice. Switch 2 selects the capacitors that determine the turnover point. The maximum boost/cut curves are arranged to shelve gently, in line with current commercial practice, rather than to continue rising or falling outside the audio range. In addition, the coupling capacitor $C_4$ has a significant impedance at 10 Hz so that the maximum bass boost curve not only shelves but begins to fall. Full boost gives +15 dB at 30 Hz but only +8 dB at 10 Hz. The tone control system has a maximum effect of ±14 dB at 50 Hz and ±12.5 dB at 10 kHz.

The tone-control amplifier uses the same low distortion configuration as the normalization stage, but it is used in a virtual-earth mode. The main difference is that the open-loop gain has been traded for open-loop linearity by increasing the emitter resistor of the main voltage amplifier from 1 to 10 kΩ thus increasing local feedback. Resistor $R_4$ has been increased to 5.6 k to maintain appropriate d.c. conditions. This modification makes it much easier to compensate for stability in the unity-gain condition that occurs when treble-cut is applied.

## Level detection circuitry

From the tone-control section the signal is fed to the final volume control via the muting reed-relay. Note that this arrangement allows the volume control to load the input of the external power amplifier even when the relay contacts are open, thus minimising noise. The signal level leaving the tone-control stage is comprehensively monitored by the circuitry shown

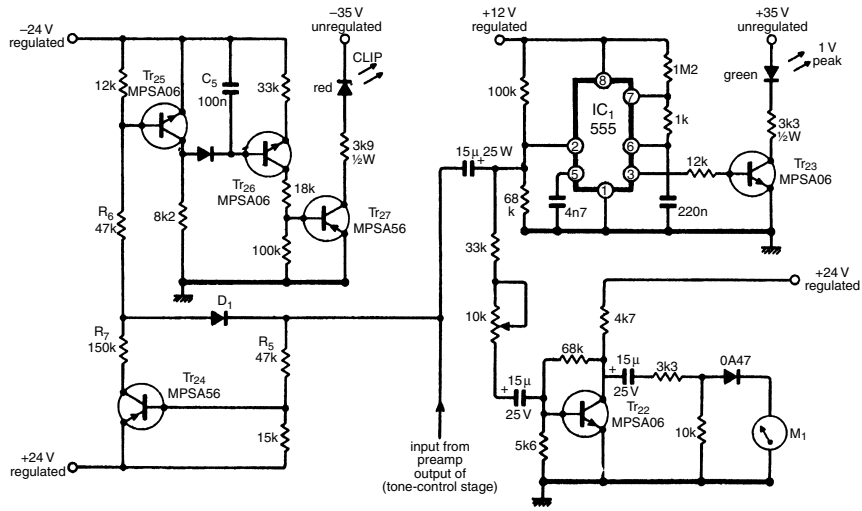



**Figure 4** *Level monitoring circuitry. Although three separate circuits are shown, these may be omitted as required.*

in Figure 4. Each channel is provided with two peak-detection systems, one lights a green l.e.d. for a pre-determined period if the signal level exceeds 1 V peak, and the other lights a red l.e.d. if the tone-control stage is on the verge of clipping. Each channel is also provided with a VU meter driver circuit. Transistor $Tr_{22}$ forms a simple amplifying stage which also acts as a buffer. Voltage feedback is used to ensure a low-impedance drive for the meter circuitry. The first peak detector is formed by $IC_1$ and its associated components. When the voltage at pin 2 goes negative of its quiescent level by 1 V, the timer is triggered and the l.e.d. turns on for a defined time. The relatively heavy l.e.d. current is drawn from an unstabilized supply to avoid inducing transients into any of the stabilized supplies.

The clipping detector continuously monitors the difference in voltage between the tone-control amplifier output and both supply rails. If the instantaneous voltage approaches either rail, this information is held in a peak-storage system. Normally $Tr_{24}$ and $Tr_{25}$ conduct continuously but if the junction of $D_1$ and $R_5$ approaches the +24 V rail then $Tr_{24}$ and hence $Tr_{25}$ turn off. This allows $C_5$ to charge and turn on $Tr_{26}$, and $Tr_{27}$ and hence the l.e.d. until the charge on $C_5$ has been drained off through emitter-follower $Tr_{26}$. If the measured voltage nears the −24 V rail, then $D_1$ conducts to pull up the junction of $R_6$ and $R_7$, which once again turns off $Tr_{25}$. In this way both positive and negative approaches to clipping are indicated. This comprehensive level indication does of course add

significantly to the task of building and testing the preamplifier. If desired, any or all of the three sections may be omitted.

## Noise gate

The final section controls the muting reed-relay. At switch-on, the +12V rail rises rapidly until stabilized by the zener diode. Pin 2 on $IC_4$ is, however, briefly held low by $C_6$, and the 555 is therefore immediately triggered to send pin 3 high. This saturates $Tr_{28}$ which prevents $Tr_{29}$ from turning on. At the end of the time delay, pin 3 goes low and relay driver $Tr_{29}$ is no longer disabled (Figure 5).

The noise gate uses two amplifiers with gains of about 100. These sample both channels at the output of the normalization stage and the inputs are clamped with diodes so that the normalization amplifiers may use their full voltage swing capability without damaging the 741s. Due to their high gain, under normal signal conditions the op-amp outputs move continuously between positive and negative saturation which keeps the storage capacitor $C_7$ fully charged. In the silent passages between 1.p. tracks the 1.f. signal is not normally of sufficient amplitude to cause saturation but will usually produce at least +3 to +4V across $C_7$ which gives a large margin of safety

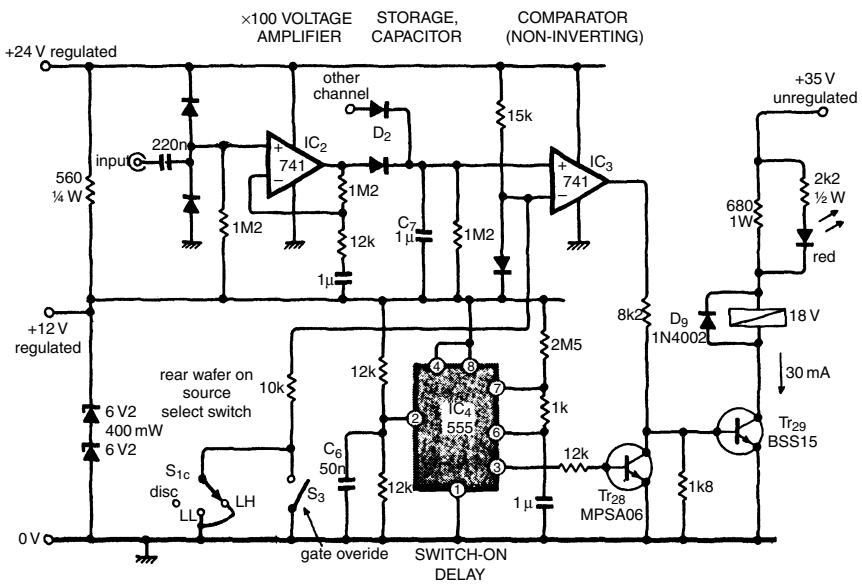



**Figure 5** *Noise gate and delay switch on circuitry. The noise gate is provided with an override switch for use with line input signals. The delay switch-on overrides all of the circuitry. Amplifier $IC_2$ is repeated for a stereo system.*

against unwanted muting. To facilitate this the response of the amplifiers is deliberately extended below the audio band. When the stylus leaves the record surface and the 1.f. signals cease, $C_7$ slowly discharges until the non-inverting input of comparator $IC_3$ falls below the voltage set on the inverting input. At this point the 741 switches and its output goes low to cut off the base drive to $Tr_{29}$, and switch off the relay. When the stylus is replaced on a record, the process takes place in reverse, the main difference being that $C_7$ charges at once due to the low forward impedance of $D_2$. To prevent the relay sporadically operating when the preamplifier is handling signals presented through the line inputs, an extra wafer on the source-select switch is arranged to override the rumble-sensing circuit, and provide permanent unmute. This is achieved by pulling the inverting input of comparator $IC_3$ negative of the +15 V rail by the 10 kΩ resistor so that even when $C_7$ is fully discharged, $IC_3$ will not switch. In addition, $S_3$ provides a manual override for testing and comparison purposes.

The power supply is shown in Figure 6. Regulators are used to provide stabilized ±24 V rails. The unregulated supply rests at about ±35 V. The signal circuitry has been designed to withstand ±35 V appearing on the supply rails, so that even in the unlikely event of both regulators failing, no further destruction will arise. Each regulator IC requires about 7 $cm^2$ of heat sink area.

Physical layout of the preamplifier is no more critical than that of any other piece of audio equipment. In general it is wise to use a layout that places the disc input amplifier as close as possible to its input socket, and as far as possible from the mains transformer. Screened cable should be

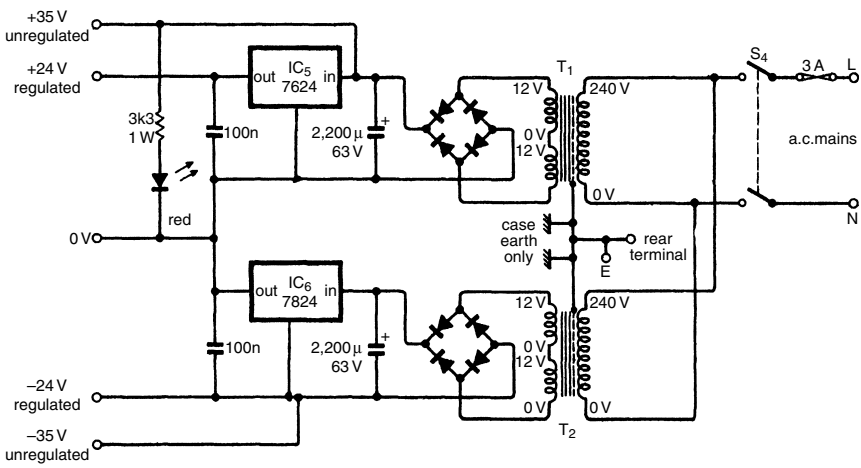



**Figure 6** *Power supply. Two regulator i.cs are used which should be mounted on heat sinks.*

used between the disc input stage and its input socket, and between the final volume control and the output socket. The earthing requirements are straightforward and the circuit common 0 V rail is led from the input sockets through the signal path to the output volume control, and finally to the 0 V terminal of the power supply. This arrangement minimises the possibility of spurious e.m.fs arising between stages. The only problem likely to be encountered is the formation of an earth loop when the preamplifier is connected to a power amplifier. Therefore, it may be satisfactory in a permanent installation to have the preamplifier circuitry connected to mains earth only through the signal lead to the power amplifier. The preamplifier case must of course be connected to the mains earth for safety reasons. It is preferable to define the potential of the preamplifier even if the power amplifier is disconnected. In the prototype the 0 V rail was connected to the mains earth via a 22 Ω resistor which stops the formation of an earth loop and prevents the signal circuitry from taking up a potential above earth due to leakage currents etc.

Testing is relatively straightforward, providing the preamplifier is constructed and checked stage by stage. Dynamic parameters such as THD are not accurately measurable without expensive test gear, but it has been found in the course of experimentation that if the d.c. conditions are correct then the various signal stages almost always show the desired a.c. performance. The non-signal circuitry should be relatively simple to fault-find. No problems should be encountered with the noise gate section which has proved to be very reliable throughout a protracted period of

## Component notes

All unmarked diodes are 1N914 or equivalent.
Red bias I.e.ds are TIL209 or equivalent.
Green bias l.e.ds are TIL211 or equivalent.
Resistors marked with an asterisk should be metal oxide types
$Tr_1$ to $Tr_6$ and $Tr_{13}$ to $Tr_{15}$ are BCY71 $Tr_{7,8,9,16,17,18,22,23,25,26,28}$ are MPS A06.
$Tr_{10,11,12,19,20,21,24,27}$ are MPS A56
$Tr_9$ is BFX85 or equivalent.
The muting reed relay should be a two pole make type with an 18 V coil.
If a different coil voltage is used, the value of the dropper resistor should be adjusted.
The VU meter should have a 1 mA movement.
If an internal diode and series resistor are fitted, the external components should be omitted.
Switch 1 (source select) is a five pole 3 way.
Switch 2 (treble frequency) is a four pole 3 way.

testing. The only preamplifier adjustment is for the VU meter calibration. This should be set to IV r.m.s. = 0 VU, which is completely non-standard but very useful in terms of the dynamic range of the signal path. For normal operation the input gain controls should be set so that the meter indications do not exceed 0 VU, to preserve a safety margin in the later stages. This completes the preamplifier design.

## References

1. Walker, H.P. 'Low-noise Audio Amplifiers', *Wireless World*, May 1972.
2. King, Gordon J. *The Audio Handbook*, Newnes-Butterworth, 1975.
3. Heidenstrom, P.N. 'Amplifier Overload', *Hi-Fi News*, December 1974.

# 2 High-performance preamplifier

## Low-cost design with active gain control

*February 1979*

This was my first 'conventional' preamplifier design to be published. It was my own reaction to the relative complexity of the Advanced Preamplifier just described; I set out to produce a preamplifier that was conventional, to see just how good it could be. At that time the available op-amps were looked at with entirely justified suspicion; they were relatively noisy and prone to crossover distortion in their output stages. Crossover might be inescapable in a power amplifier, but it was definitely not a good thing to have in a preamp. Hence the use of discrete Class-A circuitry throughout. (The 5534 op-amp was just becoming available at the time, but was ferociously expensive.)

The basic philosophy was the use of simple two or three-transistor stages, enhanced with current-source outputs when required, running from a rather high rail voltage to increase headroom and reduce distortion at a given signal level. A single supply rail was used, without regulation, but with an extra RC filter to reduce ripple to about 50 mV. This minimal-cost arrangement gave hum and noise figures as good as those from the dual-rail, IC regulator method. Where it fell down was that there was of course no DC regulation, so when the rather low noise and distortion were being measured, the audio analyser residual signal heaved up and down like a rough sea, making measurements rather tricky, even when a high-pass filter was employed.

Note the transistor equivalent of the White cathode-follower at the disc-stage output, giving push-pull Class-A operation with beautiful simplicity. Not so conventional after all.

Some years ago Doug Self described a no-compromise preamplifier which was designed using high voltage transistors to give exceptional performance. This new design sacrifices very little of that performance and uses a small number of low-cost transistors to significantly reduce the cost. A novel active gain control makes best use of the dynamic range and removes the problem of volume control placement.

This preamplifier offers a similar performance to that of the advanced preamplifier published previously,[1] but with a simpler design that reduces the parts count and hence cost. In normal use, the signal levels are kept around 50 mV by exchanging the normal potentiometer volume control, which acts as an attenuation control, for an active gain control. Therefore, the signal receives only the amplification required for a given output and so makes best use of the amplifier's dynamic range.

The distortion performance is also improved because unwanted gain will be used to give higher negative feedback and thus greater linearity. The active gain control uses a shunt feedback circuit where the volume control varies the resistance of a feedback arm as shown in Figure 1. The disc input stage has a relatively low gain of +20 dB at 1 kHz which allows a very high input overload margin. This is followed by a third-order high-pass filter which removes subsonic signals while they are still at a low level. Both bass boost and treble cut portions of the RIAA equalisation take place in the first stage. The gain control stage is positioned after the input switching and has a maximum voltage gain of +20 dB. This is followed by a Baxandall tone control which has unity gain at 1 kHz.

The use of an active gain control eliminates the problems associated with a normal volume control. If all of the gain is placed before the control, the supply voltage limits the overload margin. If some gain occurs after the volume control, then the signal-to-noise ratio is degraded because noise generated in the later stages does not undergo attenuation. The use of two controls, one early and one late in the signal chain, is one method of avoiding this compromise[1] but a true gain control is considered to be a more elegant solution.

Because a low-cost, single unregulated power supply is used with first-order RC smoothing to reduce ripple, all sections of the preamplifier are designed with high ripple-rejection performance.

## Disc input stage

The most difficult stage to design in a preamplifier is the disc input, and the problems are compounded if, as in this case, the gain of the stage is low to allow a high overload margin. A low voltage gain at 1 kHz means that the feedback network which defines the gain and RIAA equalisation will have a relatively low impedance, and thus appear as a heavy load to be driven

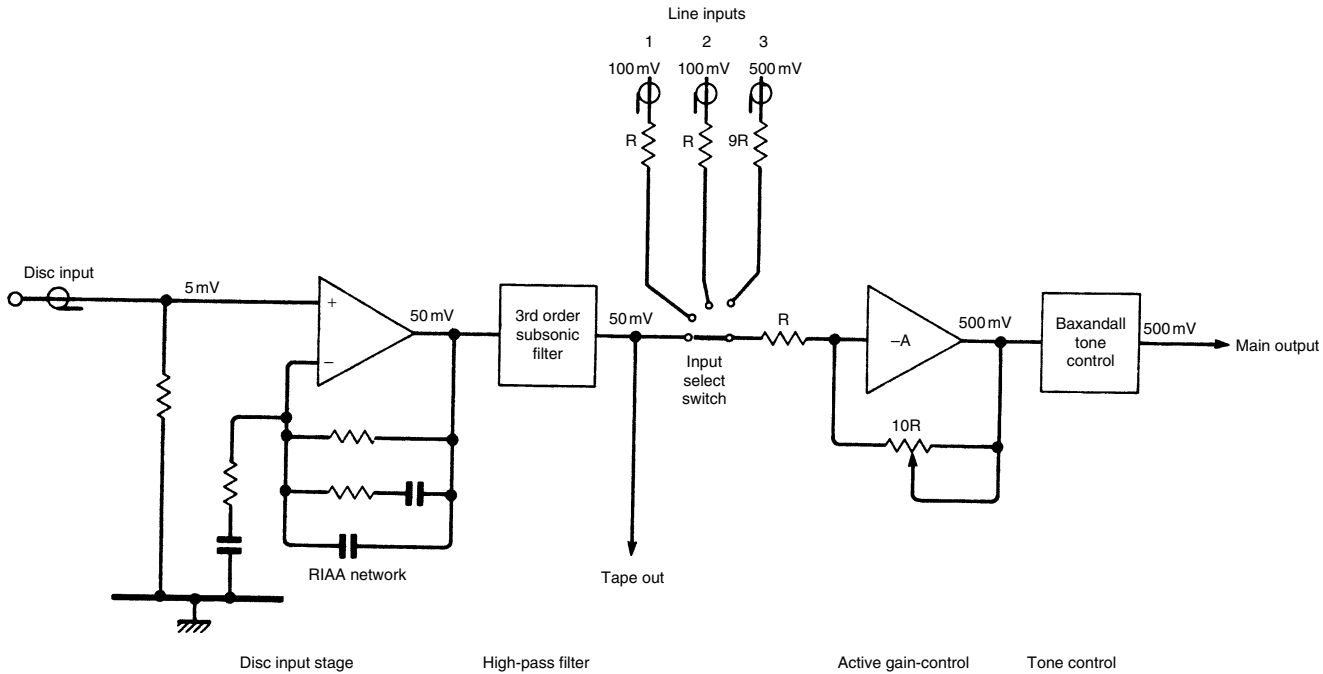



**Figure 1** *Block diagram of the preamplifier. The signal voltages shown are for maximum gain at* 1 kHz.

by the disc amplifier. This situation becomes worse at higher frequencies when the reactance of the equalisation components falls. Therefore, as a large voltage swing at the output is desirable, a large amount of current must be able to flow into and out of the feedback network at high frequencies. A second, and related problem, is that if the gain at 1 kHz is low, the gain at 20 kHz must be 19.3 dB lower due to the RIAA equalisation, which makes it close to unity. Therefore, it becomes more difficult to set the top end of the RIAA curve accurately. For this reason an extra low-pass section, with a −3 dB frequency of about 22 kHz, is added after the disc amplifier to ensure that the high-frequency gain continues to fall at a steady rate. It should be noted that if the correct turn-over frequency is chosen for the final low-pass network, the RIAA amplitude and phase curves are obtained exactly.

## Preamplifier specification

### *Input sensitivity*

for 500 mV output
Disc 5 mV at 47 kΩ
Line 1 100 mV at 20 kΩ
Line 2 100 mV at 20 kΩ
Line 3 500 mV at 100 kΩ

### *Disc input overload level*

1.1 V r.m.s. at 1 kHz
3.8 V r.m.s. at 10 kHz

### *Outputs*

Main output
500 mV r.m.s. at 100 Ω source impedance
Tape output
50 mV r.m.s. at 1 kΩ source impedance
Maximum possible level from main output
8.5 V r.m.s.

### *Frequency response*

Disc input (R1AA equalisation)
±1.0 dB 20 Hz to 20 kHz
±0.5 dB 100 Hz to 20 kHz

Line inputs (flat)
+0, to −0.5 dB 20 Hz to 20 kHz

### *Total harmonic distortion*

From disc input to main output, at 1 kHz with the gain control set to ×6
less than .008% at 8 V r.m.s.
less than .005% at 5 V r.m.s.
Because the output signal level will normally be around 500 mV the THD level will be much lower.

### *Noise*

Disc input better than 68 dB below 5 mV r.m.s.
Line inputs below −75 dBm at full gain
Residual below −90 dBm at zero gain

### *Tone controls*

Bass ±14 dB at 50 Hz
Treble ±10 dB at 10 kHz

### *Power consumption*

Approx. 78 mA each channel from a +38 V supply.

Another consequence of the fall in closed-loop gain at high frequencies is that the compensation for Nyquist stability is more difficult, and in this design it was necessary to add a conventional RC step-network to the normal dominant-pole compensation. The dominant-pole capacitor is kept as small as possible to preserve the slew-rate capability of the stage.

The basic disc input stage is shown in Figure 2. In this series-feedback configuration almost all of the voltage gain is provided by the second transistor, which has a bootstrapped collector load for high open-loop gain and linearity. The final transistor is an emitter-follower for unity-gain voltage buffering. This configuration allows the use of a p-n-p input transistor for optimum noise performance, but it also means that the collector current must flow through the feedback resistance $R_F$. This places another constraint on the design of the feedback network because an excessive voltage drop must be avoided.

As the disc input amplifier must be capable of sourcing or sinking large peak values of current to drive the capacitive feedback loop at high frequencies, the conventional emitter-following output circuit in Figure 2 is not suitable because the sink current causes a voltage drop in $R_E$. Lowering

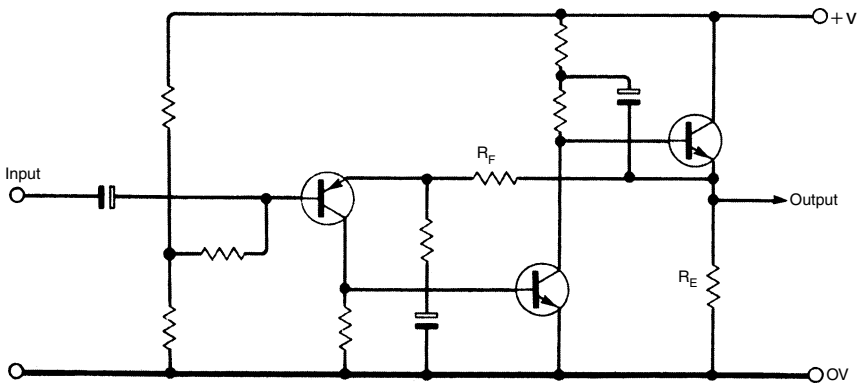



**Figure 2** *Series-feedback disc input amplifier with an emitter-follower output.*

the value of $R_E$ reduces the effect, but this is a poor solution as it leads to a high quiescent power dissipation. Replacing $R_E$ with a constant-current source is more effective because the maximum sink current becomes equal to the standing current of the stage. However, this would still limit the output of the disc stage at high audio frequencies due to an inability to sink sufficient current. For this reason, the push-pull class A configuration in Figure 3 was chosen. The bottom transistor is a current-source which is modulated in anti-phase to the top emitter-follower, via the current-sensing resistor $R_A$ and a capacitor. This can also be considered as a negative-feedback loop that attempts to keep the current in $R_A$ constant. However, the open-loop gain is only unity and so with 100% negative feedback the current variations in the top transistor are reduced to one half by the capacitor. Due to the anti-phase drive of the lower transistor, this stage can sink a peak current of twice the standing current, and therefore give twice the output swing at high frequencies.

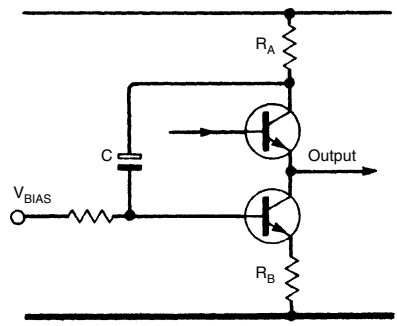



**Figure 3** *Improved push–pull class A output stage.*

A practical circuit of the disc input amplifier and its associated subsonic filter is shown in Figure 4. All of the d.c. bias voltages are provided by the potential divider $R_2$, $R_3$, $R_4$, $D_1$ and $D_2$. This chain is heavily decoupled by $C_2$ to prevent supply-rail ripple entering this sensitive part of the circuit. Note that $Tr_3$ and $Tr_5$ are isolated from the bias voltage by $R_{14}$ and $R_{22}$ to simplify any fault-finding.

The RIAA equalisation is provided by $R_{10}$, $R_{11}$, $C_6$ and $C_7$ in the feedback loop, and $R_7$, $C_4$ forms a step network that aids h.f. stability. Resistor $R_{17}$ and $C_{11}$ make up the low-pass section that corrects the top octave of the RIAA curve. The subsonic filter is a 3-pole Butterworth type with an ultimate slope of 18 dB/octave. Although the frequency response shows a loss of only 1.5 dB at 20 Hz, the attenuation is increased to more than 14 dB at 10 Hz. The unity-voltage gain element of the filter is formed by $Tr_5$ and $Tr_6$ arranged as an emitter-follower with a current-source load. This configuration was chosen for its excellent linearity. An output of about 50 mV is available for tape recording although the exact voltage will depend on the cartridge sensitivity. Resistor $R_{24}$ prevents damage to $Tr_6$ if the tape output is shorted to earth, and resistor $R_{25}$ maintains the output of the disc stage at 0 V d.c., and also prevents switching clicks.

The total harmonic distortion from input to tape output at 6 V r.m.s. is below 0.004% from 1 to 10 kHz but because the anticipated signal level here from most cartridges is about 50 mV r.m.s. The distortion during use will be even lower. The disc input will accept more than 1 V r.m.s. at 1 kHz, and about 3.8 V r.m.s. at 10 kHz before overloading. It is felt that the improvement these figures show over conventional methods justifies the complication of a low-gain disc input stage. The accuracy of the RIAA equalisation depends on how closely the RC time-constants can be set. If 5% components are used the deviation should be less than ±0.5 dB from 1 to 15 kHz, and within ±1 dB from 20 Hz to 20 kHz. The signal-to-noise ratio for a 5 mV r.m.s. input at 1 kHz is better than 68 dB.

The remaining part of the preamplifier comprises an active gain-control and the tone-control stage. The input switching is simple and requires only one switch section per channel. Also, any line input of suitable sensitivity can be used as a tape monitor return.

The shunt-feedback configuration of the active gain control enables each line input to have its sensitivity defined by the value of a single series input resistor. The maximum voltage gain available from the stage is the ratio of the feedback resistance to the input resistance, and is +20 dB when the volume control is at maximum resistance. This gain is only used in the disc mode. The most sensitive line input is rated at 100 mV for a 500 mV output and the least sensitive input has unity gain. Any sensitivity between these two limits may be provided by using the appropriate series resistor value.

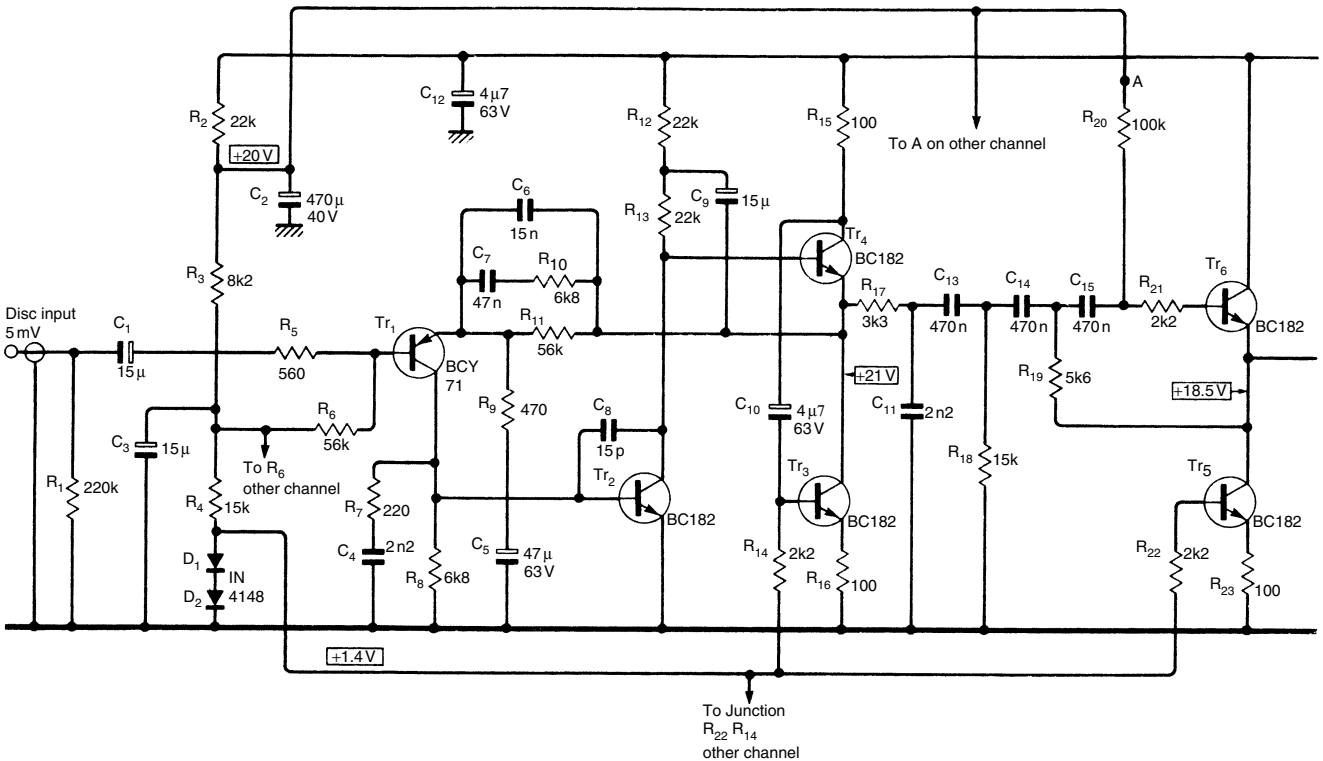



**Figure 4a** *One complete channel of the preamplifier. Components in the bias suppy, $R_2$, $R_3$, $R_4$, $C_2$, $C_3$, $D_1$, $D_2$ and also $C_{12}$ are not repeated in channel 2. All electrolytic capacitors are rated at* 40 V *and all resistors are* ¼W *unless otherwise stated.*

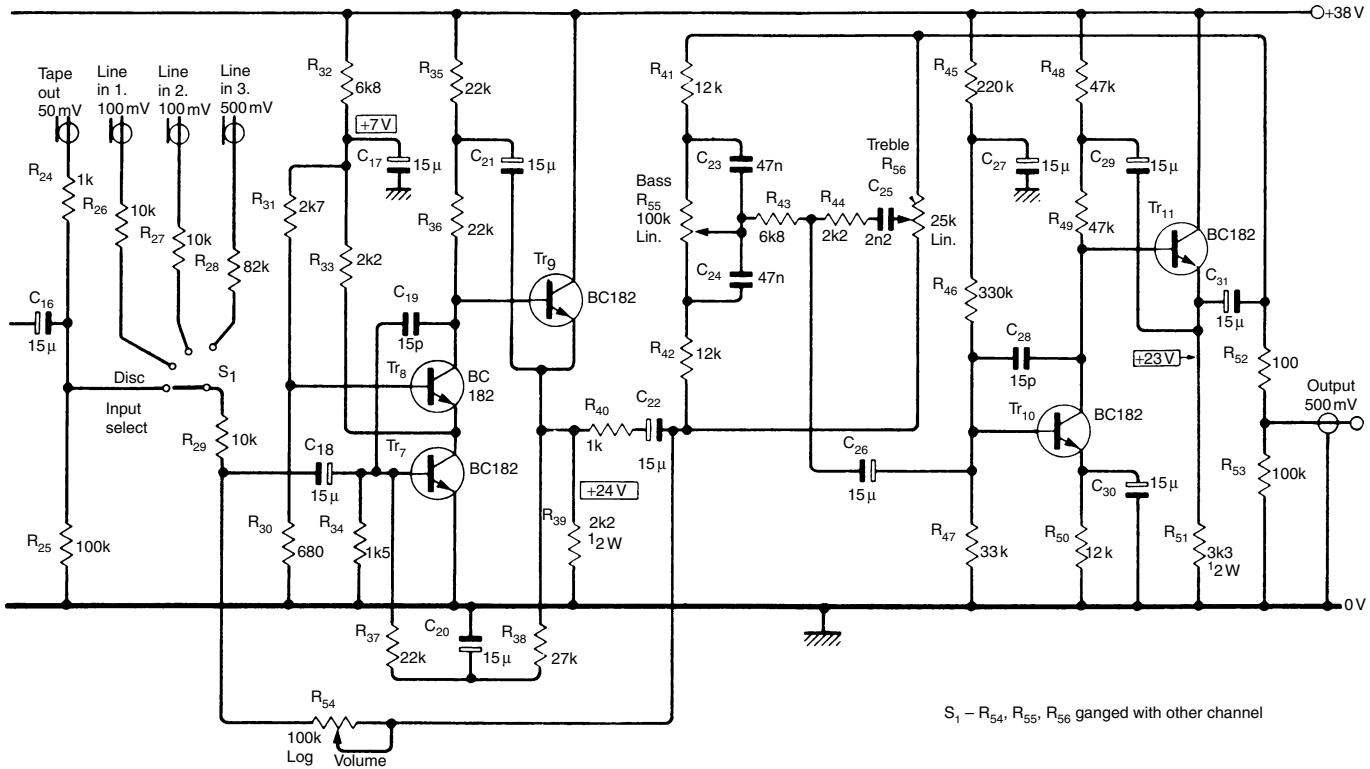



**Figure 4b** *One complete channel of the preamplifier. Components in the bias suppy, $R_2$, $R_3$, $R_4$, $C_2$, $C_3$, $D_1$, $D_2$ and also $C_{12}$ are not repeated in channel 2. All electrolytic capacitors are rated at* 40 V *and all resistors are* ¼W *unless otherwise stated.*

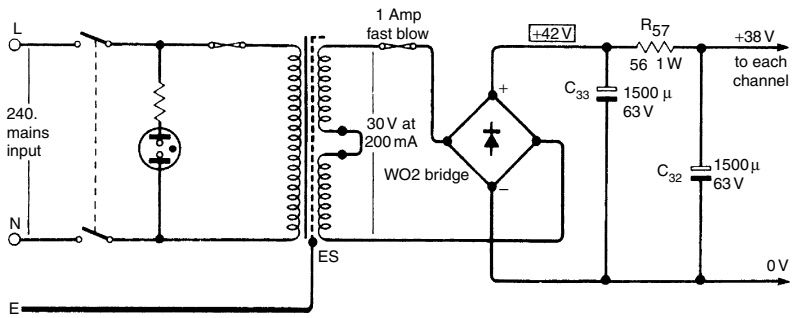



**Figure 5** *Power supply. If a higher voltage transformer is used, $R_{57}$ should be increased accordingly.*

The gain control comprises $Tr_7$ and $Tr_8$ arranged as a cascode voltage amplifier with a bootstrapped collector load and $Tr_9$ as a conventional emitter-follower. The d.c. conditions are set by negative feedback through $R_{37}$ and $R_{38}$, and a.c. feedback is applied through the volume control. The linearity of this circuit is increased by a current injected into $Tr_7$ through $R_{33}$. The voltage at the top of $R_{33}$ and the potential divider $R_{30}$, $R_{31}$, is smoothed by $R_{32}$ and $C_{17}$. Resistor $R_{40}$ prevents high-frequency instability when the volume control is set to zero gain.

The tone-control is a conventional Baxandall circuit, with $Tr_{10}$ providing a high voltage-gain by its bootstrapped collector load. Transistor $Tr_{11}$ is another emitter-follower which buffers the high impedance at the collector of $Tr_{10}$. The output is taken through $R_{52}$, which protects the output against short-circuits. Because the output impedance is low, long cables may be used without loss of high frequencies. The power supply is shown in Figure 5.

## Construction

Normal precautions should be taken to keep a.c. power away from the disc input stage, and to avoid earth loops. The leads to $R_{54}$ should be kept short to prevent hum pick-up on the virtual-earth point of the gain control. Typical voltages for various parts of the circuit are shown in Figure 4. These measurements should be made with a 20 kΩ/V meter.

Several modifications can be made to the preamplifier to suit individual requirements. Firstly, the treble turnover frequency of the tone-control section can be increased from 2 kHz as shown in Figure 4, to 5 kHz, for example, by reducing $C_{25}$ to 1000 pF. For variable turnover frequencies $C_{25}$ can be made switchable. Some purists may feel that the provision of a tone-control is unnecessary, and even undesirable. In this case, the output

should be taken from the junction of $C_{22}$ and $R_{54}$, but $R_{52}$ and $R_{53}$ should be retained at the output. Because the current drawn by the preamplifier will now be less, it is advisable to raise the value of $R_{57}$ to keep the supply rail at +38V.

In the circuit of Figure 4, no balance control is included. This function was performed in the prototype by a dual-concentric volume control. If, however, a conventional balance network is required this can be added at the output of the preamplifier although the low output impedance will be sacrificed.

## Reference

1. Self, D. 'Advanced preamplifier', *Wireless World*, November 1976, p 41.

# 3 Precision preamplifier

*October 1983*

By the time the Precision Preamplifier of 1983 was conceived, the remarkable 5534/5532 op-amps were available at a reasonable price, and delivered very low noise with almost unmeasurable distortion, given a few simple precautions. It became clear that even if ultimate performance was the goal, it was no longer economical or sensible to assemble eight or more transistors into a home-made discrete op-amp. The adoption of op-amps meant a return to dual regulated power supplies, but time and progress had made this option less costly than it had been when the Advanced Preamplifier was designed.

I put a lot of effort into producing the best design I could, taking the 'precision' bit very seriously. I found that getting the RIAA equalization accurate beyond a certain point by cut-and-try methods was virtually impossible, and very quickly exhausted my limited supplies of patience, so I fired up the BBC Model-B (computers only took a second or so to boot in those days – we have come a long way since then) and wrote some software to optimize the RIAA component selection, evaluated dozens of different volume-control laws, and minimised noise at every point; it was very well received, so it was worth it. As with several of the designs described here, I still use the prototype on a regular basis.

Until relatively recently, any audio preamplifier with pretensions to above-average quality had to be built from discrete transistors rather than integrated circuits. The 741 series of op-amps was out of the question for serious audio design, due to slew-rate and other problems, and the TL071/72 types, though in many ways excellent, were still significantly noisier than discrete circuitry. In an article some years ago,[1] I attempted to show that it was still feasible to better the performance of such devices by using simple two or three-transistor configurations.

The appearance of the 5534 low-noise op-amp at a reasonable price, has changed this. It is now difficult or impossible to design a discrete stage that has the performance of the 5534 without quite unacceptable complexity. The major exception to this statement is the design of low-impedance low-noise stages such as electronically-balanced microphone inputs or moving-coil head amplifiers, where special devices are used at the input end.

5534 op-amps are now available from several sources, in a conventional 8-pin d.i.l. format. This version is internally compensated for gains of three or more, but requires a small external capacitor (5–15 pF) for unity-gain stability. The 5532 is a very convenient package of two 5534s in one 8-pin device with internal unity-gain compensation, as there are no spare pins.

The 5534/2 is a low-distortion, low-noise device, and a typical audio stage could be expected to generate less than 0.005% THD over the range 1–20 kHz, leaving the residual distortion lost in the noise of all but the most expensive analysers. Noise performance obviously depends partly on external factors, such as source resistance and measurement bandwidth, but as an example consider the moving-magnet disc input stage shown in Figure 3. When prototyped with a TL071, the noise (with a 1 k resistor input load) was −69 dB with reference to a 5 mV r.m.s. 1 kHz input. Substituting a 5534 improved this to −84 dB, a clear superiority of 15 dB.

Another advantage of this device to the audio designer is its ability to drive low-impedance loads (down to 500 Ω in practice) to a full voltage swing, while maintaining low distortion. This property is much appreciated by studio mixer designers, whose output amplifiers are still expected to drive largely fictitious 600 Ω loads. As a comparison, the TL071 is only good for loads down to about 2 kΩ.

## Architecture

As explained in a previous article,[1] the most difficult compromise in preamplifier design is the distribution of the required gain (usually at least 40 dB) before and after the volume control. The more gain before the volume control, the lower the headroom available to handle unexpectedly large signals. The more gain after, the more the noise performance deteriorates at low volume settings. Another constraint is that it is desirable to get the signal level up to about 100 mV r.m.s. before reaching the volume control, as tape inputs and outputs must be placed before this. The only really practical way to get the best of both worlds is to use an active gain-control stage – an amplifier that can be smoothly varied in gain from effectively zero up to the required maximum.

If the input to the disc stage is a nominal 5 mV r.m.s. (assumed to be at 1 kHz throughout the avoid confusion due to RIAA equalization) from either moving-magnet cartridge or moving-coil head amp, then 26 dB of gain will be needed to give the 100 mV which is the minimum it is desirable to offer as a tape output. This can easily be got from a single 5534 stage, and taken together with the supply rails (±15 V) this immediately fixes the disc input overload at about 320 mV r.m.s. A figure such as this is quite adequate, and surpasses most commercial equipment.

One must next decide how large an output is needed at maximum volume for the 5 mV nominal input. 1 V r.m.s. is usually ample, but to be certain of being able to drive exotic units to their limits, 2 V r.m.s. is safer. This decision is made easier because using an active gain-control frees us from the fear of having excessive gain permanently amplifying its own noise after the volume control. Raising the 100 mV to this level requires the active gain stage to have another 26 dB of gain available; see the block diagram in Figure 1.

The final step in fixing the preamp, architecture is to place the tone-control in the optimum position in the chain. Like most Baxandall stages, this requires a low-impedance drive if the response curves are to be predictable, and so placing it after the active gain-control block (which has the usual very low output impedance) looks superficially attractive. However, further examination shows that (a) the active-gain stage also requires a low-impedance drive, so we are not saving a buffer stage after all, and (b) since it uses shunt feedback the tone-control stage is rather noisier than the others,[2] and should therefore be placed before the gain control so that its noise can be attenuated along with the signal at normal volume

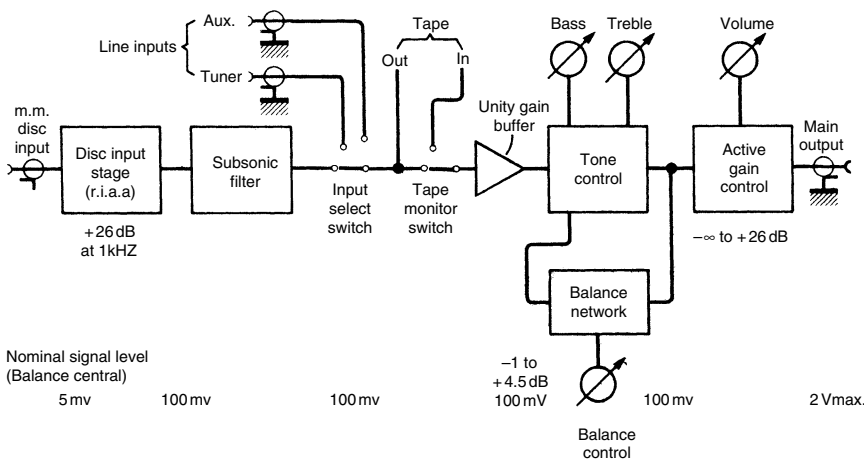



**Figure 1** *Block diagram. Tone-control placed before gain-control block to reduce noise from tone-control.*

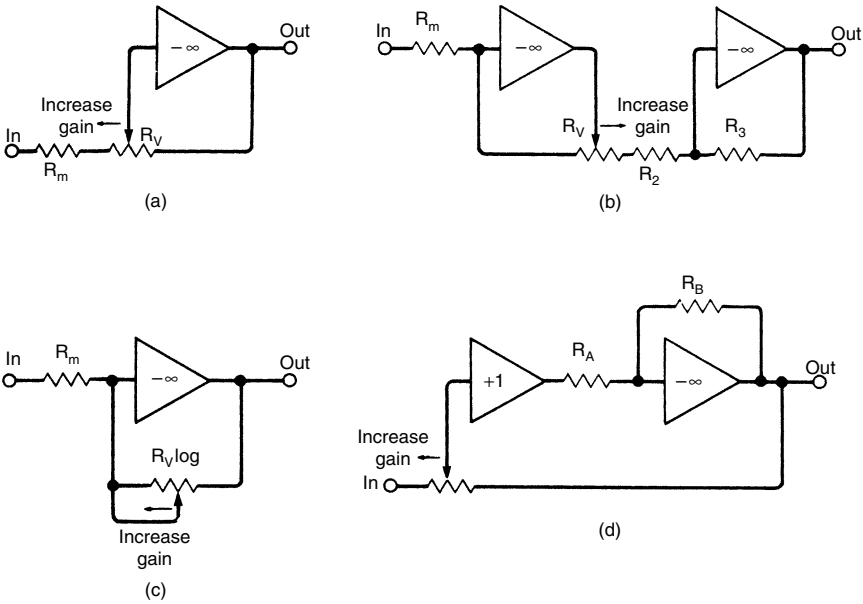
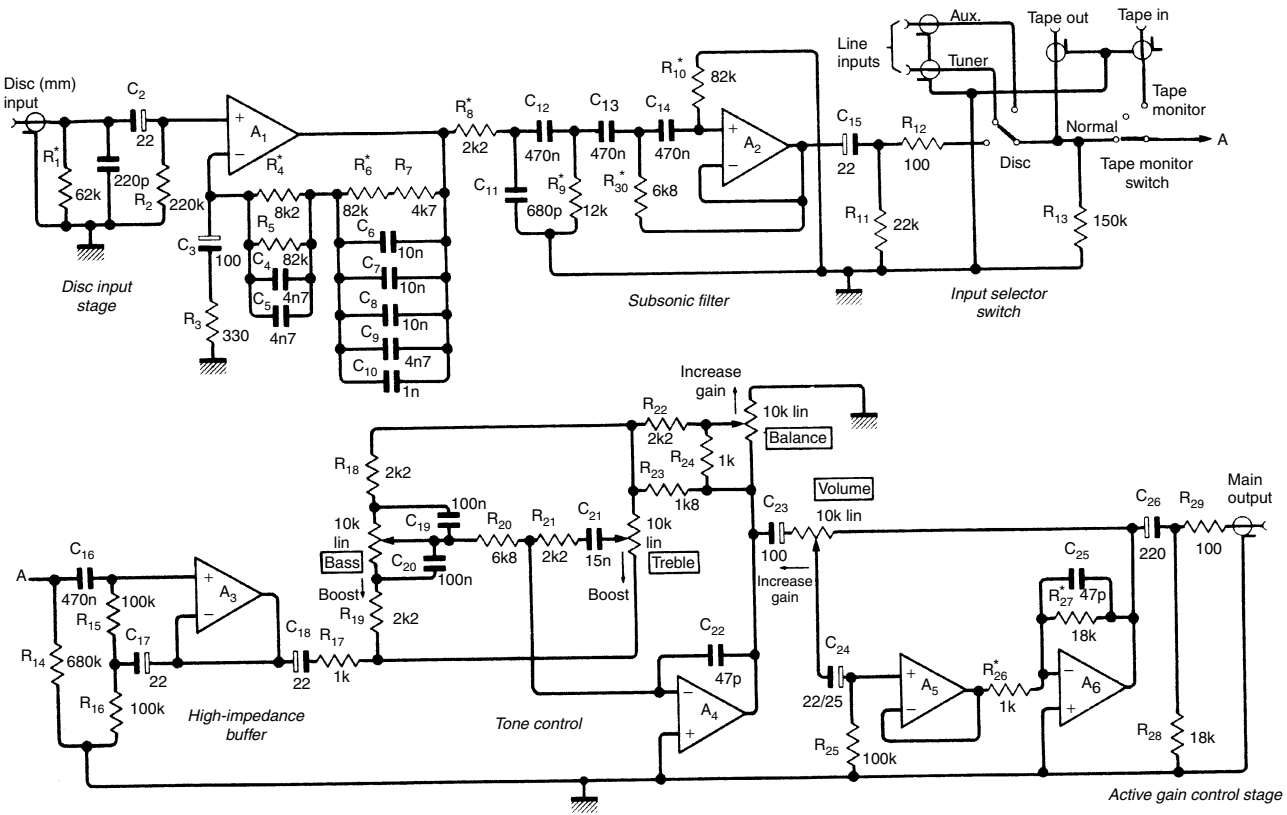



**Figure 2** *Evolution of active gain-control stage. That due to Baxandall, chosen for this design, is at (d).*

settings. The tone-control is preceded by a unity-gain buffer stage with low output impedance and a very high input impedance, so that the load placed on line input devices does not vary significantly when the tape-monitor switch is operated. This brings us to the block diagram in Figure 1. Figure 3 shows the circuit diagram of the complete preamplifier. The components around $A_1$ and $A_2$ make up the moving-magnet disc stage and its associated subsonic filter. Disc preamplifier stage $A_1$ uses a quite conventional series feedback arrangement to define the gain and provide RIAA equalisation. This provides a clear noise-performance advantage of 13 dB over the shunt feedback equivalent,[2,3] which is sometimes advocated on the rather dubious grounds of 'improved transient response'. The reality behind this rather woolly phrase is that the series configuration cannot give the continuously descending frequency response in the ultrasonic region that the RIAA specification seems to imply, because its minimum gain is unity. Hence sooner or later, as the frequency increases, the gain levels out at unity instead of dropping down towards zero at 6 dB per octave. As described in Refs. 1 and 2, when a low-gain input stage is used to obtain a high overload margin, 'sooner' means within the audio band, and so an additional low-pass time-constant is required to cancel out the unwanted h.f. breakpoint; once more it is necessary to point out that if the low-pass time-constant is correctly chosen, no extra phase or amplitude errors are


Disc (mm) input
C2
22
R1*
62k
220p
R2
220k
A1
R4*
R6*
R7
82k
4k7
R5
8k2
C3
100
C4
82k
C5
4n7
R3
330
4n7
C6
10n
C7
10n
C8
10n
C9
4n7
C10
1n
Disc input stage
R8*
2k2
C12
470n
C13
470n
C14
470n
R10*
82k
C11
680p
R9*
12k
R30*
6k8
A2
Subsonic filter
Line inputs
Aux.
Tuner
Tape out
Tape in
Tape monitor
C15
22
R12
100
Disc
Normal
A
Tape monitor switch
R11
22k
R13
150k
Input selector switch
Increase gain
10k lin
Balance
R22
2k2
R24
1k
R23
1k8
R18
2k2
C19
100n
10k lin
Bass
C20
100n
R20
6k8
R21
2k2
C21
15n
10k lin
Treble
Boost
Boost
R19
2k2
C23
100
Volume
10k lin
Increase gain
C26
220
R29
100
Main output
C25
R27*
47p
18k
C16
470n
A
R15
100k
C17
22
R14
680k
R16
100k
A3
C18
22
R17
1k
High-impedance buffer
C22
47p
A4
Tone control
C24
22/25
R25
100k
A5
R26*
1k
A6
R28
18k
Active gain control stage

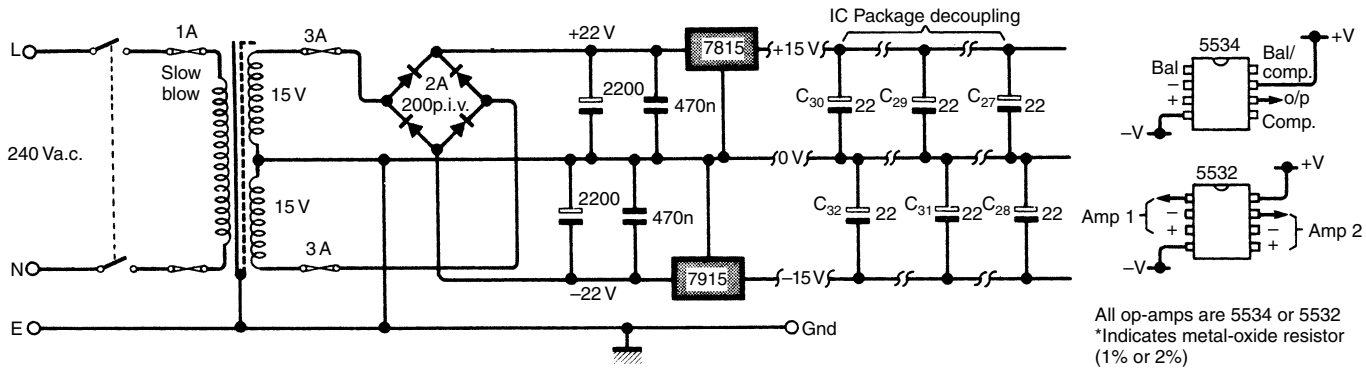



**Figure 3** *Complete circuit diagram. Decoupling capacitors for i.cs must be close to packages.*

introduced. This function is performed in Figure 3 by $R_8$ and $C_{11}$, which also filter out unwanted ultrasonic rubbish from the cartridge.

It was intended from the outset to make the RIAA network as accurate as possible, but since the measuring system used (Sound Technology 1700 A) has a nominal accuracy of 0.1 dB, 0.2 dB is probably the best that could be hoped for. Designing RIAA networks to this order of accuracy is not a trivial task with this configuration, due to interaction between the time-constants, and attempting it empirically proved most unrewarding. However, Lipshitz, in an exhaustive analysis of the problem, using heroic algebra in quantities not often seen, gives exact but complicated design equations.[4] These should not be confused with the rule-of-thumb time-constants often quoted. The Lipshitz equations were manipulated on an Acorn Atom microcomputer until the desired values emerged. These proved on measurement to be within the 0.2 dB criterion, with such errors as existed being ascribable to component tolerances.

Design aims were that the gain at 1 kHz should be 26 dB, and that the value of $R_3$ should be as small as feasible to minimize its noise contribution. These two factors mean that the RIAA network has a lower impedance than usual, and here the load-driving ability of the 5534 is helpful in allowing a full output voltage swing, and hence a good overload margin.

There is a good reason why the RIAA capacitors are made up of several in parallel, when it appears that two larger ones would allow a close approach to the correct value. It is pointless to design an accurate RIAA network if the close-tolerance capacitors cannot be easily obtained, and in general they cannot. The exception to this is the well-known Suflex range, usually sold at 2.5% tolerance. These are cheap and easy to get, the only snag being that 10 nF seems to be the largest value widely available, and so some paralleling is required. This is however a good deal cheaper and easier than any other way of obtaining the desired close-tolerance capacitance.

Metal-oxide resistors are used in the RIAA network and in some other critical places. This is purely to make use of their tight tolerance (1% or 2%), as tests proved, rather unexpectedly, that there was no detectable noise advantage in using them.

The recently updated RIAA specification includes what is known as the 'IEC amendment'. This adds a further 6 dB/octave low-cut time-constant that is −3 dB at 20.02 Hz. It is intended to provide some discrimination against subsonic rumbles originating from record warps, etc., and in a design such as this, with a proper subsonic filter, it is rather redundant. Nonetheless the time-constant has been included, in order to keep the bottom octave of the RIAA accurate. The time-constant is *not* provided by $R_3C_3$ (which is no doubt what the IEC intended) but by the subsonic filter itself, a rather over-damped third-order Butterworth type designed so that its slow initial roll-off simulates the 20.02 Hz time-constant, while below 16 Hz the reponse drops very rapidly. Implementing the IEC roll-off by

reducing $C_3$ is not good enough for an accurate design due to the large tolerances of electrolytic capacitors. However, the $R_3, C_3$ combination is arranged to roll-off lower down (−3 dB at about 5 Hz) to give additional subsonic attenuation.

Capacitor $C_1$ defines the input capacitance and provides some r.f. rejection. A compromise value was chosen, and this may be freely modified to suit particular cartridges.

The noise produced by the disc input stage alone, with its input terminated with a 1 k resistor to simulate roughly a moving-magnet cartridge, is −84.5 dB with reference to a 5 mV r.m.s. 1 kHz input (i.e. 100 mV r.m.s. out) for a typical 5534 A sample. The suffix A denotes selection for low noise by the manufacturer. When the 1 k termination is replaced by a short circuit, the level drops to −86 dB, indicating that in real life the Johnson noise generated by the cartridge resistance is significant, and so that stage is really as quiet as it is sensible to make it.

## Subsonic filter

As described above, this stage not only rejects the subsonic garbage that is produced in copious amounts by even the flattest disc, but also implements the IEC roll-off. Below 16 Hz the slope increases rapidly, the attenuation typically increasing by 10 dB before 10 Hz is reached. The filter therefore gives good protection against subsonic rumbles, that tend to peak in the 4–5 Hz region.

This filter obviously affects the RIAA accuracy of the lowest octave, and so $C_{12}$, $C_{13}$, $C_{14}$ should be good-quality components. A 10% tolerance should in practice give a deviation at 20 Hz that does not exceed 0.7 dB, rapidly reducing to an insignificant level at higher frequencies. The tape output is taken from the subsonic filter, with $R_{12}$ ensuring that long capacitative cables do not cause h.f. instability. If it really is desirable to drive a 600 Ω load, then $C_{15}$ must be increased to 220 μF to maintain the base response.

## High-impedance buffer

This buffer stage is required because the following tone-control stage demands a low-impedance drive, to ensure that operating the tape monitor switch $S_2$ does not affect the tape-output level. If the input selector switch $S_1$ was set to accept an input from a medium impedance source (say 5 k), and the buffer had a relatively low input impedance (say 15 k), then every time the tape-monitor switch was operated there would be a step change in level due to the change of loading on the source. This is avoided in

this design by making the buffer input impedance very high by conventional bootstrapping of $R_{15}$, $R_{16}$ via $C_{17}$. This is so effective that the input impedance is defined only by $R_{14}$. Unlike discrete-transistor equivalents, this stage retains its good distortion performance even when fed from a high source resistance, e.g. 100 k.

## Tone-control stage

Purists may throw up their hands in horror at the inclusion of this, but it remains a very useful facility to have. The range of action is restricted to ±8 dB at 10 kHz and ±9 dB at 50 Hz, anything greater being out of the realm of hi-fi. The stage is based on the conventional Baxandall network with two slight differences. Firstly the network operates at a lower impedance level than is usual, to keep the noise as low as possible. The common values of 100 k for the bass control and 22 k for the treble control give a noise figure about 2.5 dB worse. Even with the values shown, the tone stage is about 6 dB noisier than the buffer that precedes it. Both potentiometers are 10 k linear, which allows all the preamplifier controls to be the same value, making getting them a little easier. The low network impedance also reduces the likelihood of capacitative interchannel crosstalk. Once again, implementing it is only possible because of the 5534's ability to drive low-value loads.

Secondly, the tone-control stage incorporates a vernier balance facility. This is also designed as an active gain-control, with the same benefit of avoiding even small compromises on noise and headroom. The balance control works by varying the amount of negative feedback to the Baxandall network, and therefore some careful design is needed to ensure that the source resistance of the balance section remains substantially constant as the control is altered, or the frequency response may become uneven. Resistors $R_{22}$, $R_{23}$, $R_{24}$ define this source resistance as 1 k, which is cancelled out by $R_{17}$ on the input side. The balance control has a range of +4.5 to −1.0 dB on each channel, which is more than enough to swing the stereo image completely from side to side. If you need a greater range than this, perhaps you should consider siting your speakers properly.

## Active gain-control stage

An active gain-control stage must fulfil several requirements. Firstly, the gain must be smoothly variable from maximum down to effectively zero. Secondly, the law relating control rotation and gain should be a reasonable approximation to logarithmic, for ease of use. Finally, the use of an active

stage allows various methods to be used to obtain a better stereo channel balance than the usual log. pot. offers.

All the configurations shown in Figure 2 meet the first condition, and to a large extent, the second. Figures 2(a) and 2(b) use linear controls and generate a quasi-logarithmic law by varying both the input and feedback arms of a shunt-feedback stage. The arrangement of Figure 2(c), as used in the previous article, offers simplicity but relies entirely on the accuracy of a log. pot. While 2(a) and 2(b) avoid the tolerances inherent in the fabrication of a log. track, they also have imperfect tracking of gain, as the maximum gain in each case is fixed by the ratio of a fixed resistor $R_m$ to the control track resistance, which is not usually tightly controlled. This leads to imbalance at high gain settings.

Peter Baxandall solved the problem very elegantly,[1] by the configuration in 2(d). Here the maximum gain of the stage is set not by a fixed-resistor/track-resistance ratio, but by the ratio of the two fixed resistors $R_a$, $R_b$. A buffer is required to drive $R_a$ from the pot. wiper, because in a practical circuit this tends to have a low value. It can be readily shown by simple algebra that the control track resistance now has no effect on the gain law, and hence the channel balance of such a system depends only on the mechanical alignment of the two halves of a dual linear pot. The resulting gain law is shown in Figure 4, where it can be seen that a good

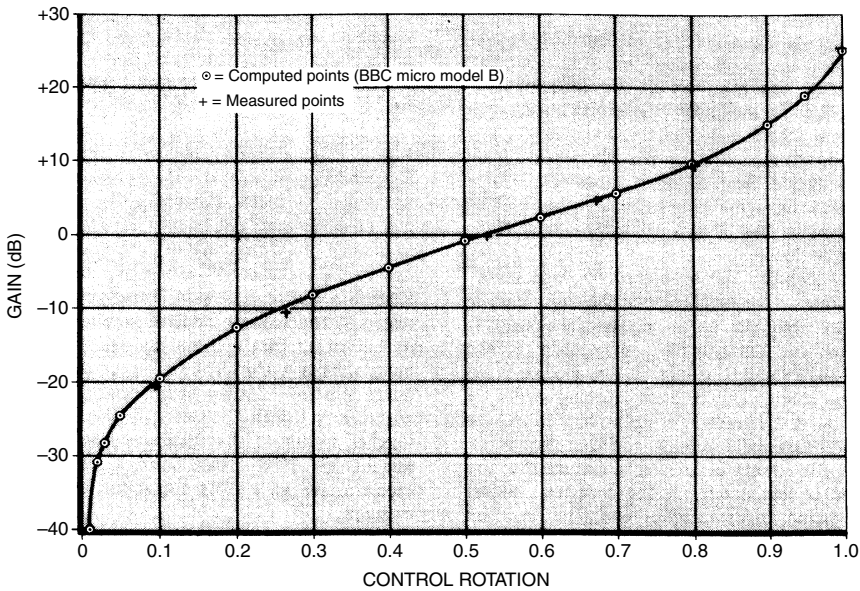



**Figure 4** *Law of gain-control pot., approximately linear over main part of range.*

approximation to the ideal log (i.e. linear in dB) law exists over the central and most used part of the control range.

A practical version of this is shown in Figure 3. $A_5$ is a unity-gain buffer biased via $R_{25}$, and $R_{26}$, $R_{27}$ set the maximum gain to the desired +26 dB. Capacitor $C_{25}$ ensures h.f. stability, and the output capacitor $C_{26}$ is chosen to allow 600 Ω loads to be driven. A number of outwardly identical Radiohm 20 mm dual-gang linear pots were tested in the volume control position, and it was found that channel balance was almost always within ±0.3 dB over the gain range −20 to +26 dB, with occasional excursions to 0.6 dB. In short, this is a good way of wringing the maximum performance from inexpensive controls, and all credit must go to Mr Baxandall for the concept.

At the time of writing there is no consensus as to whether the absolute polarity of the audio signal is subjectively important. In case it is, all the preamplifier inputs and outputs are in phase, as the inversion in the tone stage is reversed again by the active-gain stage.

## Power supply

The power supply is completely conventional, using complementary i.c. regulators to provide ±15 V. Since the total current drain (both channels) is less than 50 mA, they only require small heatsinks. A toriodal mains transformer is recommended for its low external field, but it should still be placed as far as possible from the disc input end of the preamplifier. Distance is cheaper (and usually more effective) than Mu-Metal. Since the 5534 is rated up to ±20 V supplies, it would be feasible to use ±18 V to get the last drop of extra headroom. In my view, however, the headroom already available is ample.

## Construction

The preamplifier may be built using either 5534 op-amps or the 5532 dual type. The latter are more convenient (requiring no external compensation) and usually cheaper per op-amp, but can be difficult to obtain. To compensate each 5534 for unit gain, necessary for each one, connect 15 pF between pins 5 and 8. Note that the rail decoupling capacitors should be placed as close as possible to the op-amp packages – this is one case in which it really does matter, as otherwise this i.c. type is prone to h.f. oscillation that is not visible on a scope, but which results in a very poor distortion performance. It must also be borne in mind that both the 5534 and 5532 have their inputs tied together with back-to-back parallel diodes, presumably for voltage protection, and this can make fault-finding with a voltmeter very confusing.

Only 2.5% capacitors should be used in the RIAA networks if the specified accuracy is to be obtained. Resistors in Figure 3 marked * should be metal oxide 1% or 2%, for reasons of tolerance only. Each of these resistors sets a critical parameter, such as RIAA equalization or channel balance, and no improvement, audible or otherwise, will result from using metal oxide in other positions.

Several preamplifier prototypes were built on Veroboard, the two channels in separate but parallel sections. The ground was run through in a straight line from input to output. Initially the controls were connected with unscreened wire, and even this gave acceptable crosstalk figures of about −80 dB at 10 kHz, due to the low circuit impedances. Screening the balance and volume connections improved this to −90 dB at 10 kHz, which was considered adequate. It must be appreciated that the crosstalk performance depends almost entirely on keeping the two channels physically separated.

Some enthusiasts will be anxious to (a) use gold-plated connectors; (b) by-pass all electrolytics with non-polarized types; or (c) remove all coupling capacitors altogether, in the pursuit of an undefinable musicality. Options (a) and (b) are pointless and expensive, and (c) while cheap, may be dangerous to the health of your loudspeakers. Anyone wishing to dispute these points should arm themselves with objective evidence and a stamped, addressed envelope.

## Specification

(Based on measurements made on three prototypes, with Sound Technology 1710A.)

### *Moving-magnet*

| | |
|---|---|
| noise ref. 5 mV r.m.s., 1 kHz input | −81 dB |
| RIAA accuracy | ±0.2 dB |
| input overload point (1 kHz) | 300 mV r.m.s. |

### *Line inputs*

| | |
|---|---|
| noise ref. 100 mV r.m.s. i/p | −85 dB |
| maximum input | 9 V r.m.s. |
| maximum gain | +26 dB |
| treble control range | ±8 dB |
| bass control range | ±9 dB |

| | |
|---|---|
| vernier balance control | −1 dB to +4.5 dB |
| volume control channel balance | ±0.3 dB |
| distortion (1 kHz–20 kHz) | 0.005% |
| maximum output | 9.5 V r.m.s. |

## References

1. Self, D. 'High-Performance Preamplifier', *Wireless World*, February 1979.
2. Walker, H.P. 'Low noise audio amplifiers', *Wireless World*, May 1972, pp 233–237.
3. Linsley-Hood, J. 'Modular preamplifier', *Wireless World*, October 1982, p 32 onwards.
4. Lipshitz, S.P. 'On RIAA equalisation networks', *J. Audio Eng. Soc.*, June 1979, p 458 onwards.
5. Baxandall, P. 'Audio gain controls', *Wireless World*, November 1980, pp 79–81.

# 4 Design of moving-coil head amplifiers

*December 1987*

For many years moving-coil head amps were exotic devices of highly specialised design. They were usually all-discrete, to minimise noise, and because of this, and the very low feedback impedances to be driven, linearity was poor, and only acceptable because the signal levels were so low.

This design is a hybrid discrete/op-amp configuration, which gives very low distortion even at signal levels monstrously above the normal operating conditions. It sidesteps several problems by having much more gain than is normally required; this would normally be a very bad move, severely curtailing headroom, but here it works as it can be assumed that the MC amp is always followed by a moving-magnet disc stage with substantial gain, so clipping will occur there first. When the article appeared, the 2SB737 transistor, with its magnificently low Rb, was expensive and not that easy to obtain, but in a year or two this situation changed and nobody now would consider devices like the 2N4403 for this application.

In recent years, moving-coil cartridges have increased greatly in popularity. This is not the place to try and determine if their extra cost is justified by an audible improved performance; suffice it to say that a preamplifier now needs a capable moving-coil cartridge input if it is to be considered complete. The head-amplifier design presented here as an example was originally intended to be retrofitted to the precision preamplifier previously published in *Wireless World*,[1] feeding the existing moving-magnet disc input. However, it is adaptable to almost any preamplifier and cartridge as the gain range available is very wide; it should therefore be of interest to any engineer working in this field. Hereafter 'moving coil' is abbreviated to m.c., and 'moving-magnet' to m.m.

Traditionally, moving-coil cartridges were matched to moving-magnet inputs by special transformers, which give 'free gain' – in a sense – and are capable of a good noise performance if the windings are carefully designed for very low series resistance. However, the inescapable problems of low-frequency distortion, high-frequency transient overshoots and the need for obsessive screening to avoid 50 Hz mains pickup render them unattractive and expensive.

The requirements for a high-quality m.c. head-amplifier are as follows. The overwhelming need is for a good noise performance, as the signals generated by m.c. cartridges are, in general, very low. However, this sensitivity is also much more variable than that of m.m. cartridges, where one can take a nominal output of 5 mV r.m.s. for 5 cm/s at 1 kHz as being virtually standard. In contrast, a survey of the available m.c. cartridges gave a range from 2.35 mV (Dynavector DV10X IV) to 0.03 mV (Audionote 102 vdH), though these are both exceptional and the great majority fell between 0.2 mV and 0.4 mV. Figure 1 shows the output levels of a number of current m.c. cartridges plotted on a scale of dBu (i.e. referred to 775 mV) and m.m. cartridges are included on the right for comparison. It is notable how these bunch together in a range of less than 7 dB.

A representative m.c. cartridge used both as a basis for design, and for testing, is the Ortofon MC10 Super, which has an output of 0.3 mV for 5 cm/s, and an internal resistance of 3 Ω. There is general agreement that this is a good-sounding component.

As detailed above, there is a need for easily variable gain over a wide range. This can be quite adequately provided in switched steps, avoiding the problems of uncertain stereo balance on dual potentiometers. From the above output figures, a gain range of 6 dB to 46 dB appears necessary to cater for all possible cartridges. It would seem, at the low-gain end, that the amplifier is virtually redundant, and so a minimum gain of 20 dB was chosen.

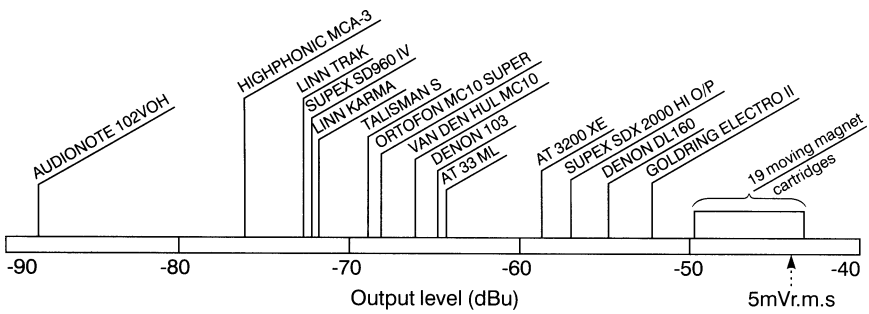



**Figure 1** *Output levels of representative moving-coil cartridges plotted on a scale of decibels relative to* 0.775 V (1 mW *in* 600 Ω)*, with the outputs of a number of moving-magnet cartridges as a comparison.*

Moving-coil cartridges are very tolerant of the loading they see at an amplifier's input, as a result of their own very low internal impedance. For example, Ortofon, who might be reckoned to know a thing or two about m.c. cartridges, simply state that the recommended load for most of their wide range of cartridges is 'greater than 10 Ω'. Nonetheless, since experimenting with cartridge loading is a harmless enough pastime, provision for changing the input loading resistor over a wide range has been made in this design.

The preamplifier should have the ability to drive a normal m.m. cartridge input at sufficient level to ensure that the head amplifier does not limit the disc headroom. Any figure here over about 300 mV r.m.s. should be satisfactory. A less obvious point is that the input impedance, apart from the nominal 47 k resistive component, usually includes a fair amount of capacitance, either to adjust cartridge frequency response or to exclude r.f. This can cause head-amplifier instability unless it is dealt with.

Finally, a head amplifier should meet the usual requirements for frequency response, crosstalk, and linearity. Capacitive crosstalk is usually not a problem, due to the very low impedances involved, but for the same reason, linearity can present problems despite the low signal levels.

## Design problems

The theoretical noise characteristics of amplifiers have been dealt with very competently in other articles,[2] and there is no need to repeat the various mathematical derivations here. The designer's options are usually limited to choosing a suitable input device, operating it at roughly the right current, (not usually critical due to the flat bottoms of the noise curves) and then making sure that the surrounding circuitry doesn't mess things up too much. M.c. head amplifiers are almost always built around discrete devices, with or without the addition of an accompanying op-amp (for an exception see Ref. 3). Figure 2 shows the reason why: when source resistances are low (say below 1 k) even advanced op-amps are easily outperformed by discrete devices, due to the inevitable compromises in i.c. fabrication. The values of equivalent input noise (e.i.n.) in Figure 2 were taken from five samples of each device, using a source resistance of 3R3, and the general circuit configuration in Figure 3. The rather non-standard measurement bandwidth is due to the use of the internal filters on a Sound Technology measuring system; adding a third-order 20 kHz active filter at the ST input would be very difficult, as the levels of noise being measured are so low. To convert to 20 kHz upper bandwidth limit, subtract 1.5 dB. One of the prerequisites for good performance in this role is a low value for $R_b$, and this has led to a fine miscellany of devices being applied to a job they were never intended for: medium power devices, print-hammer

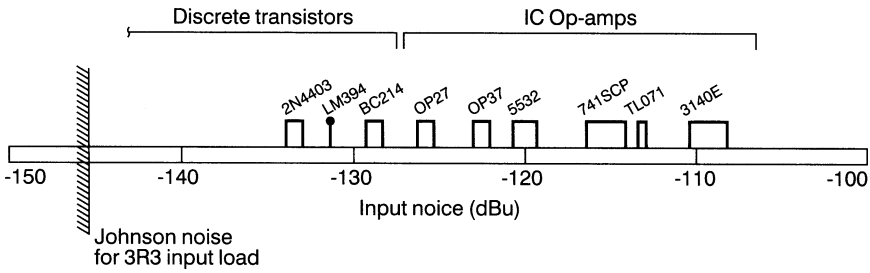



**Figure 2** *Discrete transistors still, in the main, provide a better noise performance than op-amps at low source resistances, as shown here for five examples of each type.*

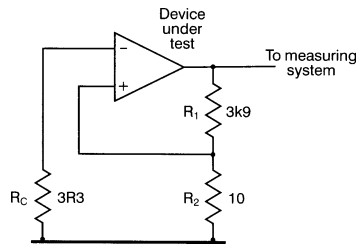



**Figure 3** *Circuit used to obtain the measurements shown in Figure 2.*

drivers, (a lot of transistors seem to have been designed as print-hammer drivers) and so on.

Apart from careful device selection, the other classical way of reducing noise with low source impedances is to use multiple devices. The assumption here is that m.c. amplifier noise will swamp the miniscule Johnson noise inherent in the source (this is usually all too true) and therefore, if two input devices have their outputs summed, the signals will simply add, giving a 6 dB gain, while the two uncorrelated device noise contributions will partially cancel, giving only 3 dB.

Thus, there is a theoretical gain of 3 dB in noise performance every time the number of input devices is doubled. There are, of course, clear economic limits to the amount of doubling you can go in for; eight parallel devices is the most that I have seen. It also seems difficult in practice to get the full theoretical benefit.

M.c. head-amplifiers in use today can be roughly divided into three common topologies, as shown in Figure 4. That shown in 4(a) relies on a single device with low $R_b$, and the combination of limited open-loop gain and the heavy loading of the low-impedance of the feedback network on the final transistor means that both linearity and maximum output level tend to be uninspiring. Given the technical resources that electronics can

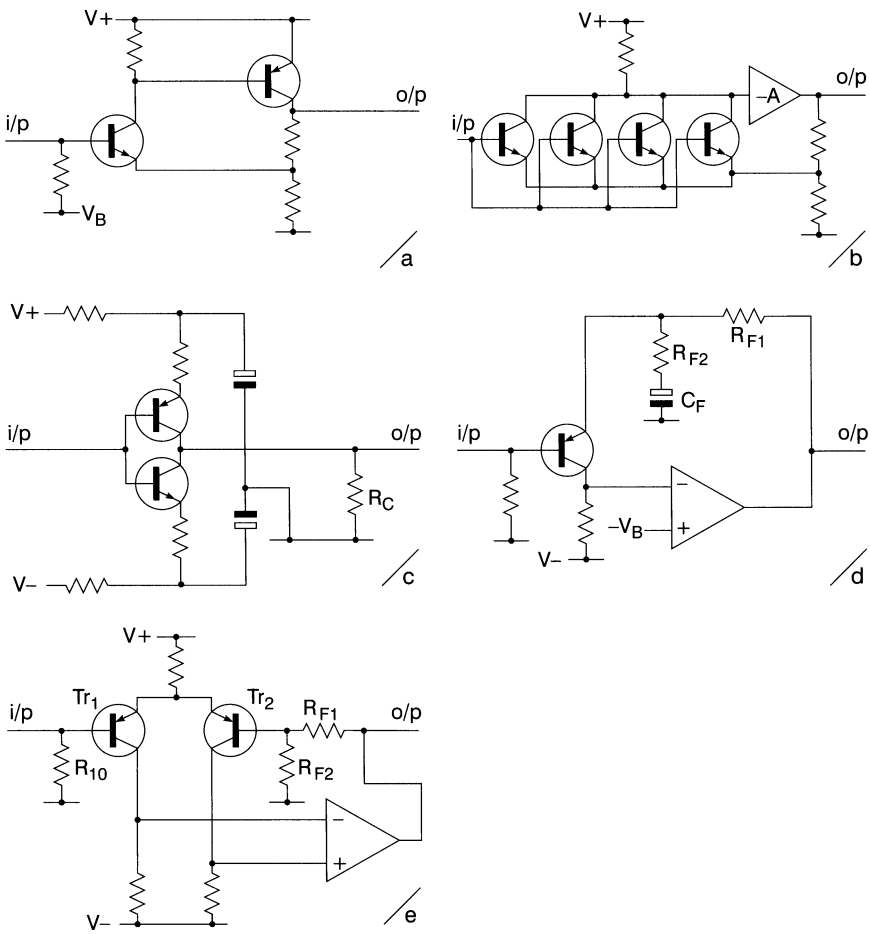



**Figure 4** *Some head-amplifier configurations. A fairly low open-loop gain in the circuit at (a) results in poor linearity. At (b), the gain is provided by multiple transistors, which theoretically gives an improvement of* 3 dB *in noise performance for twice the number of transistors, but can also present current-sharing problems. The arrangement at (c) provides the* 3 dB *improvement without current sharing: linearity is not of the highest order. Circuit (d) uses one input device, the gain being provided by an ip-amp: the necessity for Cf presents problems, which are overcome in the (e) configuration at the expense of a lowered noise performance.*

deploy, there seems no need to ask the paying customers to put up with any measurable distortion at all. An amplifier of this type is analysed in Ref. 4.

Figure 4(b) shows the classic multiple-parallel-transistor configuration; the amplifier block A is traditionally one or two discrete devices, that usually

have difficulty in driving the low-impedance feedback network. Effort is usually expended in ensuring proper current-sharing between the input devices.

This can be done by adding small emitter resistors to swamp $V_{be}$ variations, but these will effectively appear in series with the source resistance, and compromise the noise performance unless they are individually decoupled with a row of very large electrolytics. Alternatively, each transistor can be given its own d.c. feedback loop to set up its collector current, but this tends to be even more prodigal of components. Having said this, experiment proved that the problem of current-sharing was not as serious as conventional wisdom holds; this is explained below. For examples of circuitry see Ref. 5.

Figure 4(c) shows the series-pair scheme. This simple arrangement allows two input devices to give the normal 3 dB noise improvement without current-sharing problems as substantially the same collector current goes through each device. The collector signal currents are summed in $R_c$, which must be reasonably low in value to absorb-any current imbalance. This configuration has its adherents but it also has its difficulties, such as indifferent linearity.

It was therefore originally decided to base the design presented here on a single well-chosen device, with the spadework of providing open-loop gain and output drive capability left to an op-amp. This leads to the configuration in Figure 4(d), which gives excellent linearity, and less than 0.002% THD at full output may be confidently expected. The first problem to be dealt with is the very low value of $R_{f2}$; this must be as low as possible (say 10 Ω) as it is effectively in series with the input source resistance and will degrade the noise performance accordingly. This means that $C_f$ must be very large, of the order of 2200 μF, to preserve the l.f. response. A 3R3 resistor in the $R_{f2}$ position demands 4700 μF to give −3 dB at 10 Hz; this is not elegant. The capacitance $C_f$ cannot be dispensed with, since there is a d.c. level of +0.6 V on the emitter of the input device, leading to a wholly impossible offset at the output of the op-amp.

One solution to this is the use of a differential pair, as in Figure 4(e). This cancels out the $V_{be}$ of the input transistor $Tr_1$, at the cost of some degradation in the noise performance of the circuit, and hopefully the d.c. offset is so much smaller that, if $C_f$ is omitted and the offset is amplified by the full a.c. gain, it will not seriously reduce the output voltage swing. In effect, the second transistor $Tr_2$ is an emitter follower transferring the feedback signal to the emitter of $Tr_1$, and such a circuit element introduces a small but inescapable amount of extra noise. In this case, with the component values shown, the degradation is about 2.8 dB.

A possibly more serious objection to this circuit is that the offset at the output is non-negligible, about 1 V, much of which is due to the base bias current flowing through $R_n$. A d.c.-blocking capacitor on the output is

essential, and if it is an electrolytic there may be some doubt as to which way round to put it, as the exact level of input pair balance is unpredictable.

After practical trials, it was decided that a 3 dB noise penalty was too great, and that a way had to be found to use a single-ended input.

## A new approach

The new method evolved is shown in the block diagram Figure 5. There is no $C_f$ in the feedback loop, and indeed no overall d.c. feedback at all. The two halves of the circuit, the input transistor and the op-amp, each have their own d.c., feedback systems. The transistor relies on simple shunt negative feedback via d.c. loop 1, while the op-amp has its output held precisely to a d.c. level of 0 V by the integrator $A_2$. This senses the mean output level, and sets up a voltage on the non-inverting input of $A_1$ that is very close to the level set on $Tr_1$ collector, such that the output stays firmly at zero; its time-constant is made large enough to ensure that an ample amount of open-loop gain exists at the lowest audio frequencies. Failure to do this results in a rapid rise of distortion as the frequency is lowered. Any changes in the direct voltage on $Tr_1$ collector are completely uncoupled from the output. However, a.c. feedback passes through $R_{f1}$ as usual and ensures that the linearity of the compound arrangement is near-perfect, as is often the case with transistor op-amp hybrid circuits. Due to the high open-loop gain of A the a.c. level on $Tr_1$ collector is very small and so a.c. feedback through d.c. loop I does not significantly affect the input impedance of the amplifier, which is about 8 kΩ.

The device chosen for the input transistor was the 2N4403, a type that has been acknowledged as superior for low-noise applications for some years. The $R_b$ is quoted as about 40 Ω.[5,6] More modern purpose-designed

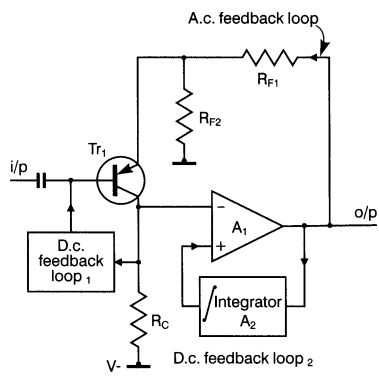



**Figure 5** *The layout adopted for the final design.*

devices such as the 2SB737 will improve the noise performance by up to 1 dB, but the extra cost is significant.

A single device used in the circuit of Figure 6 gives an e.i.n. of −138 dB with a 4 mA collector current, which is certainly not bad, but it was consistently found that putting devices in parallel without any current-sharing precautions whatever always resulted in a significant improvement in noise performance. On average, adding a second transistor reduced noise by 1.2 dB, and adding a third reduces it by another 0.5 dB. Beyond this the law of diminishing returns sets in and, since further multiplication was judged unprofitable, a triple-device input was settled on. The current-sharing under these conditions was checked by measuring the voltage across 100 Ω resistors inserted in the collector paths. Using 3.4 mA as the total current for the array it was found after much device-swapping that the worst case of imbalance was 0.97 mA in one transistor and 1.26 mA in another. No attempt was made to ensure that all the devices came from the same batch. It therefore appears that, for this device at least, matching is good enough to make simple paralleling worthwhile, and it was therefore decided to use three devices in parallel in the final circuit.

There now remains the problem of setting the gain. Usually it would be simple enough to alter $R_{f1}$ or $R_{f2}$, but here it is not quite so simple. The resistance $R_{f2}$ is not amenable to alteration, as it must be kept to the lowest practicable value of 3.3 Ω, and $R_{f1}$ must be kept up to a reasonable value so that it can be driven to a full voltage swing by an op-amp output. This means a minimum of 500 Ω if the op-amp is to be of an easily obtainable type such as the 5534. (It is paradoxical that amplifiers whose output is measured in millivolts are required to chuck around so much current.)

These two values fix a minimum closed-loop gain of about 44 dB, which is far too high for all but the most insensitive cartridges. The only solution is to use a ladder output attenuator to reduce the overall gain; this would be anathema in a conventional signal path, because of the loss of headroom involved, but since an output of 300 mV r.m.s. would be enough to overload virtually all m.m. inputs, we can afford to be prodigal with it. If the gain of the head amplifier is set to be a convenient 200 × (+46 dB) then attenuation to reduce overall gain to a more useful +20 dB still allows a maximum output of 480 mV r.m.s.; this comfortably exceeds the input capability of the intended host preamplifier, though one previous design would accept it all and come back for more.[7] Smaller degrees of attenuation to provide intermediate gains allow greater outputs, and these are summarized in the specification. The Ortofon MC10 was used with +26 dB of gain, to give similar output levels to m.m. cartridges driving the precision preamplifier RIAA stage direct.

The last constraint is the need to provide a low output impedance to the succeeding m.m. input stage, so that it can give a good noise performance;

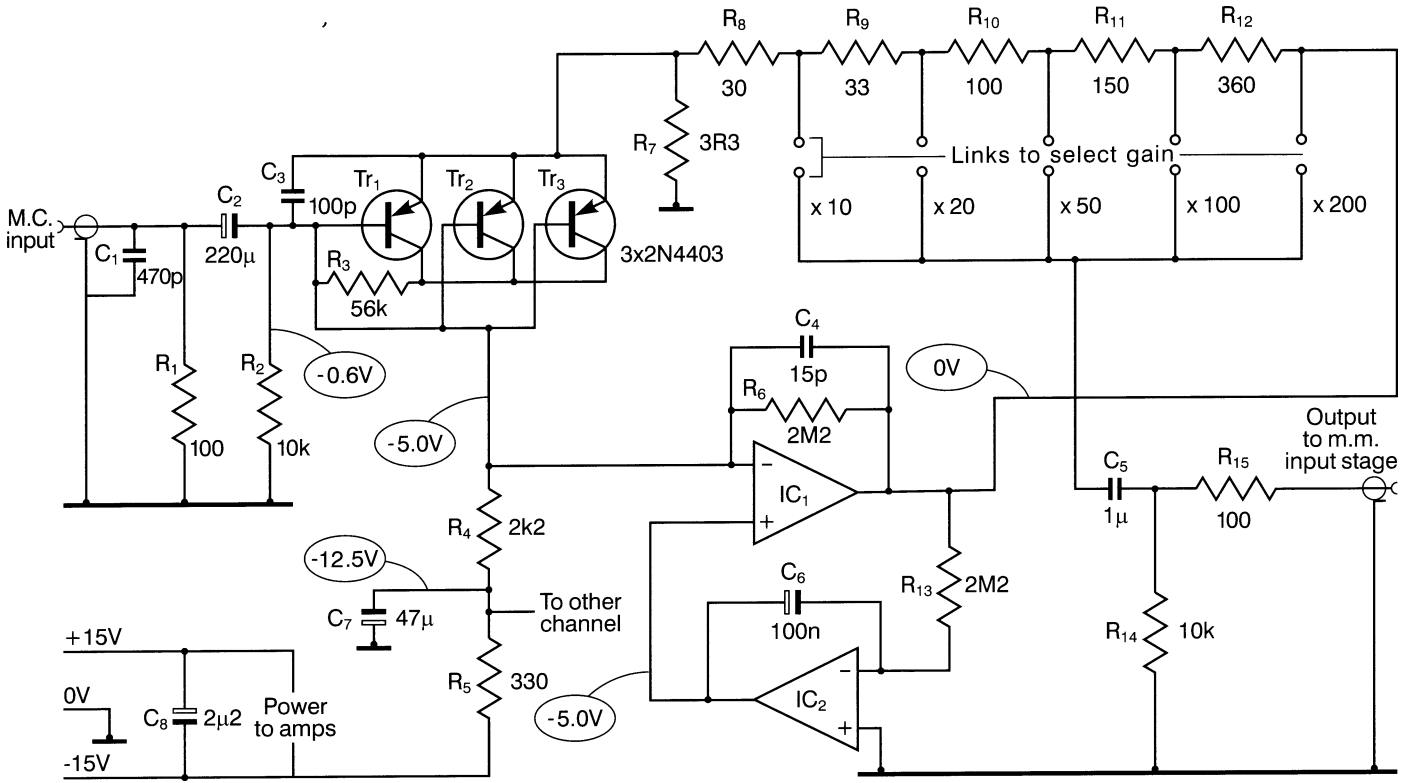



**Figure 6** *Complete circuit diagram of the moving-coil head amplifier, intended to drive the moving-magnet input of a preamplifier.*

it is likely to have been optimized to give of its best with a source impedance of 500 Ω or less. This implies that the ladder attenuator will need low resistor values, imposing yet more loading on the unfortunate op-amp, so this problem has been side-stepped by making the ladder an integral part of the a.c. feedback loop, as shown in Figure 6. This is only practicable because it is known that the load resistance presented by the next stage will be too high at 47 kΩ to cause any significant gain variations.

## The final circuit

This is shown in Figure 6, and most closely follows the configuration of Figure 4(d), with the exception that the input devices have suddenly multiplied themselves by three. Capacitor $C_1$ is soldered on the back of the m.c. input phono sockets and is intended for r.f. filtering rather than modification of the cartridge response. If the need for more capacitive or resistive loading is felt, then extra components may be freely connected in parallel with $R_1$. If $R_1$ is raised in value, then load resistances of up to 5 kΩ are possible, as the impedance looking into $C_2$ is about 8 kΩ. Capacitor $C_2$ is large to give the input devices the full benefit of the low source impedance, and its value should not be altered. Resistors $R_2$, $R_3$ make up d.c. loop 1, setting the d.c. operating conditions of $Tr_{1,2,3}$, while $R_4$ is the collector load, decoupled from the supply rail by $C_9$ and $R_5$, which are shared between the two channels. Opamp $IC_1$ provides the main a.c. open-loop gain, and is stabilized at h.f. by $C_4$ : $R_6$ has no real effect on normal operation, but is included to give $IC_1$ a modicum of negative feedback and hence tidy behaviour at power-up, when this would otherwise be lacking due to the charging time of $C_2$ the other op-amp, $IC_2$, is the integrator that makes up d.c. loop 2, its time-constant carefully chosen to provide plenty of open-loop gain from $IC_1$ at low frequencies, and to avoid a peaking in the l.f. response that can occur due to the second time-constant of $C_2$.

The ladder resistors $R_8$–$R_{12}$ make up the combined feedback-network and output-divider, overall gain being selected by a push-on link in the prototype. A rotary switch could be used instead, but this will produce loud clicks when moved with the volume up, since the emitter current of $Tr_1$–$Tr_3$ flows through $R_7$, and a small current therefore flows down the divider chain. The output resistor $R_{15}$ ensures stability when driving long screened cables, and $C_5$ is included to eliminate any trace of d.c. offset from the output because the stage might find itself driving a horribly vulnerable 'esoteric' input stage with direct coupling and possibly substantial gain at d.c. Anything is possible these days.

## Comparing performance parameters

These are given in the specification, and I think there will be few opportunities to quibble. On the vital question of noise it would be instructive to compare it with other preamplifiers – not easy because the noise performance of m.c. head amplifiers is specified in so many different ways it is virtually impossible to reduce them all to a similar form, particularly without knowing the spectral distribution of the noise. Noise performances are specified with and without CCIR-ARM weighting, over different band-widths, and with different source impedances. This article has dealt throughout with unweighted noise referred to the input, over a 400 Hz–30 kHz bandwidth, and with RIAA equalisation *not* taken into account. Without getting bogged down in invidious comparisons, I can only say that it is my belief that the design given here is quieter than most current designs, being within 6 dB of the theoretical minimum.

When using this design with the precision preamplifier, it was noted with some surprise that it was so quiet that the m.m. RIAA stage actually caused the noise performance to deteriorate by about 3 dB. Since the RIAA stage is itself very quiet (s/n ratio −81 dB referred to 5 mV r.m.s. input) it is considered that the design goals were met.

## Specification

Careful earthing is needed if the noise and cross talk performance quoted is to be obtained.

| Gain | Gain (dB) | Max output (r.m.s.) |
|---|---|---|
| 10× | +20 dB | 480 mV |
| 20× | +26 dB | 960 mV |
| 50× | +34 dB | 2.4 V |
| 100× | +40 dB | 4.6 V |
| 200× | +46 dB | 10 V |

**Input overload level.** 48 mV r.m.s.
**Equivalent input noise.** −139.5 dBu, unweighted, no RIAA.
**THD** Less than 0.002% at 7 V r.m.s. output, (maximum gain) at 1 kHz. Less than 0.004% 40 Hz–20 kHz.
**Frequency response.** +0, −2 dB, 20 Hz–20 kHz.
**Crostalk.** Less than −90 dB 1 kHz–20 kHz (layout dependent).
**Power requirements.** 20 mA at ±15 V, for both channels.

## Practice

P.c.b. layouts of require some care if the full performance is to be realised. First, the grounding should be carefully planned, as it must be realised that with such low impedances as $R_7$ (3R3) playing a vital role, the resistance of tracks can be significant. It is suggested that a single star ground point be chosen on the p.c.b., and critical paths (input ground, $R_1$, $R_7$) all connected to this, to prevent signal currents causing voltage drops where they are least wanted. It is vital to avoid making loops in the input path that will pick up 50 Hz magnetic fields.

It is essential to place the decoupling capacitor $C_8$ next to $IC_1$ to prevent insidious h.f. oscillation which makes its presence known only by severely impaired linearity. When interfacing the head amplifier to an existing design, note that about 8 mA flows down the ground connection.

## References

1. Self, D. 'A precision preamplifier', *Wireless World*, October 1983, pp 31–34.
2. Walker, H.P. 'Low-noise audio amplifiers', *Wireless World*, May 1972.
3. Duncan, B. 'AMP-01', *HiFi News*, October 1984, pp 67–73.
4. Nordholt and Van Vierzen, 'Ultra Low Noise Preamp For Moving-Coil Phono Cartridges', *JAES*, April 1980, pp 219–223.
5. Barleycorn, J. (a.k.a. S. Curtis) *HiFi For Pleasure*, August 1978, pp 105–106.
6. Foord, A. 'Introduction to low-noise amplifier design', *Wireless World*, April 1981, p 71.
7. Self, D. 'High-quality preamplifier', *Wireless World*, February 1979, p 40.

# 5 Precision preamplifier '96, Part I

*July/August 1996*

Sometimes you just feel that the world needs another preamplifier. Having put a lot of effort into power amplifiers and their problems in the meantime, I thought it would be interesting to see if the original Precision Preamplifier (Chapter 3) could be significantly improved.

The 5532-based moving-magnet disc stage was retained with a few changes (most importantly a cost-effective way to tighten up the RIAA accuracy) as attempts to make a hybrid stage that was quieter were not encouraging; a matter I still feel I have not got quite to the bottom of. The latter stages were radically altered to improve input selection crosstalk and upgrade the tone-control section by giving it variable turnover frequencies. A lot of work went into using linear pots everywhere, with suitable control-laws obtained by various devious means; the reasoning behind this was not so much to make ordering the parts easier (though that was an incidental benefit), but to avoid the channel imbalances inherent in log and anti-log pots. The variable-frequency tone-controls worked very well indeed, although this approach went dead against the prevailing fashion for omitting tone-controls altogether; this did not bother me greatly.

A new preamp design is timely. There is more variation in audio equipment than ever before, so to a greater extent preamps are required to be all things to all persons. High source resistance outputs and low-impedance inputs must be catered for, as well as ill-considered and exotic cabling with excessive shunt capacitance. The last preamp design I placed before the public was in 1983,[1] extended in facilities by the moving-coil head amp stage published in 1987.[2]

In the last ten years, small-signal analogue electronics has undergone few changes. Most circuitry is still made from *TL072s*, with resort to 5532s when noise and drive capability are important. In this period many new op-amps have appeared, but few have had any impact on audio design; this is largely a chicken/egg problem, for until they are used in large numbers the price will not come down low enough for them to be used in large numbers. Significant advantage over the old faithfuls is required.

This new design uses the architecture established in Ref. 1, which has not been improved upon so far. The already low noise levels have been further reduced. The tone controls were fixed-frequency, and proved inflexible compared with the switched-turnover versions in my previous designs,[3,4] so these frequencies are now fully variable, and a non-interrupting tone-cancel facility provided.

This preamplifier is designed to my usual philosophy of making it work as well as possible, by the considered choice of circuit configurations etc., rather than the alternative approach of specifying exotic components and hoping for the best.

## Adding tape facilities and tone control

There are two basic architectures for tape record/replay handling. The simpler, in Figure 1(d), adds a tape output and a tape monitor switch for off-tape monitoring on triple-head machines.

The more complex version in Figure 1(e) allows any input to be listened to while any input is being recorded, though how many people actually do this is rather doubtful. This method demands very high standards of crosstalk inside the preamp. There is usually no tape return input or tape monitor switch as there is now no guarantee that the main path signal comes from the same original source as the tape output.

The final step is to add tone controls. They need a low-impedance drive for predictable equalisation curves, and a vital point is that most types – including the Baxandall – phase-invert. Since the maintenance of absolute polarity is required, this inversion can conveniently be undone by the active gain control, which also uses shunt feedback and phase-inverts. The tone-control can be placed before or after the volume control, but if afterwards it generates noise that cannot be turned down. Putting it before the volume control reduces headroom if boost is in use, but since maximum boost is only +10 dB, the preamp inputs will not overload before 3 V r.m.s is applied; domestic equipment can rarely generate such levels. Figure 1(f) shows the final architecture.

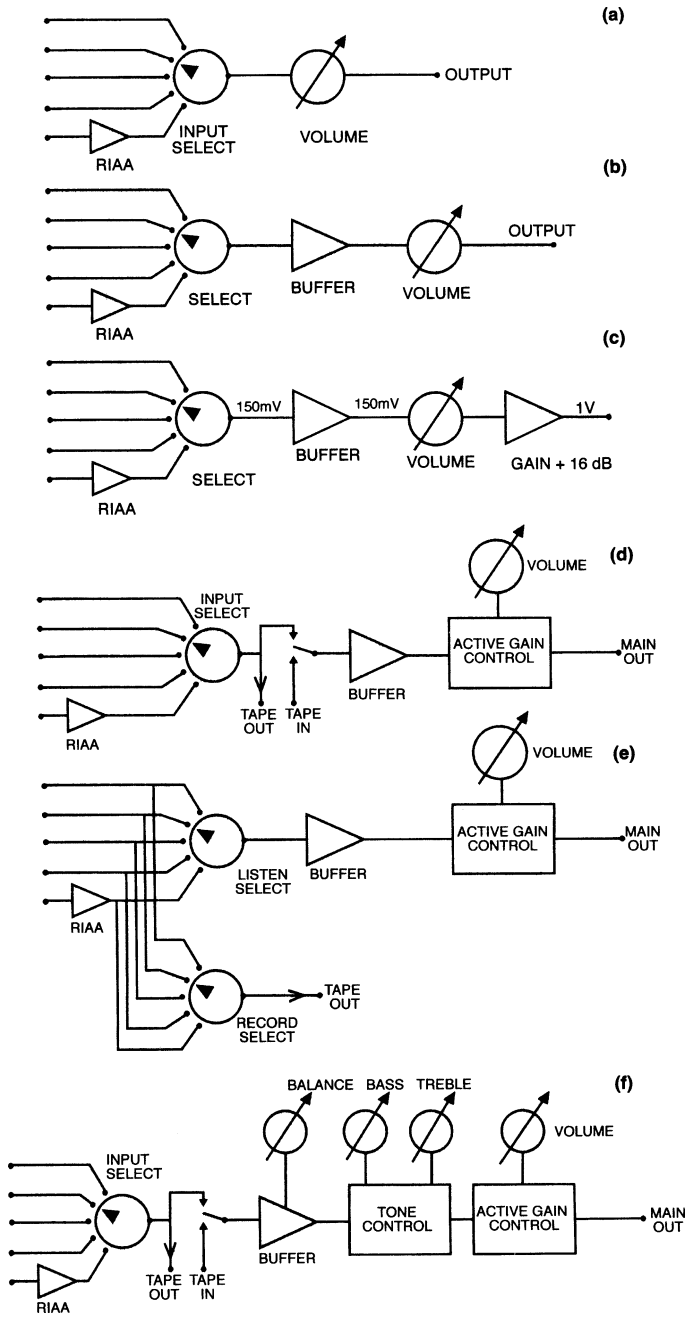



**Figure 1** *The course of preamp evolution, as impedance and level matching problems are dealt with.*

## The evolution of preamplifiers

Minimal requirements are source selection and level control, as in Figure 1(a); an RIAA disc preamp stage is one input option. This sort of 'passive preamplifier' (a nice oxymoron) is only practical if the main music source is a low-impedance high-level output like CD.

The only parameter to decide is the resistance of the volume pot; it cannot be too high because the output impedance, which reaches a maximum of one quarter the track resistance at −6 dB, will cause high-frequency roll-off with the cable capacitance. On the other hand, if the pot resistance is too low, the source equipment will be unduly loaded. If the source is valve equipment, which does not respond well to even moderate loading, the problem starts to look insoluble.

Adding a unity-gain buffer stage after the selector switch, Figure 1(b), means the volume control can be reduced to 10 kΩ, without loading the sources. This still gives a maximal output impedance of 2.5 kΩ, which allows you only 5.4 m of 300 pF/m cable before the response is 1.0 dB down at 20 kHz. For 0.1 dB down at 20 kHz, only 1.6 m is permissible.

The input *RC* filters found on so many power-amps as a gesture against transient intermodulation distortion add extra shunt capacitance ranging from 100 pF to 1000 pF, and can cause additional unwanted h.f. roll off.

Unfortunately only a CD source can fully drive a power amplifier. Output levels for tuners, phono amps and domestic tape machines are of the order of 150 mV rms, while power amplifiers rarely have sensitivities lower than 500 mV. Both output impedance and level problems are solved by adding a second amplifier stage as Figure 1(c), this time with gain. The output level can be increased and the output impedance kept down to 100 Ω or lower.

This amplifier stage introduces its own difficulties. Nominal output level must be at least 1 V r.m.s. (for 150 mV in) to drive most power amps, so a gain of 16.5 dB is needed. If you increase the full-gain output level to 2 V r.m.s., to be sure of driving exotica to its limits, this becomes 22.5 dB, amplifying the input noise of the gain stage at all volume settings. Noise performance thus deteriorates markedly at low volume levels – the ones most of us use most of the time.

One answer is to split the gain before and after the volume control, so that there is less gain amplifying the internal noise. This inevitably reduces headroom before the volume control. Another solution is double gain controls – an input-gain control to set the internal level appropriately, then an output volume control that requires no gain after it.

Input gain controls can be separate for each channel, doubling as a balance facility.[3] However this makes operation rather awkward. No matter

how attenuation and fixed amplification are arranged, there are going to be trade-offs on noise and headroom.

All compromise is avoided by an active gain stage, i.e. an amplifier stage whose gain is variable from near-zero to the required maximum. You get lower noise at gain settings below maximum, and the ability to generate a quasilogarithmic law from a linear pot. This gives excellent channel balance as it depends only on mechanical alignment.

## Design philosophy

There is great freedom of design in small-signal circuitry, compared with the intractable problems of power amplification. Hence there is little excuse for a preamp that is not virtually transparent, with very low noise, crosstalk and THD.

### Requirements for the RIAA network

- The RIAA network must use series feedback, as shunt feedback is 14 dB noisier.
- Correct gain at 1 kHz. Sounds elementary, but you try calculating it.
- Accuracy. The 1983 model was designed for ±0.2 dB accuracy 20–20 kHz, which was the limit of the test gear I had access to at the time. This is tightened to ±0.05 dB without using rare parts.
- It must use obtainable components. Resistors will be E24 series and capacitors E12 at best, so intermediate values must be made by series or parallel combinations.
- $R_o$ (Figure 2), must be as low as possible as its Johnson noise is effectively in series with the input signal. This is most important in moving-coil mode.
- The feedback network impedance to be driven must not be low enough to increase distortion or limit output swing – especially at high frequencies.
- The resistive path through the feedback arm should ideally have the same d.c. resistance as input bias resistor $R_{18}$ (Figure 8), to minimise offsets at A1 output.

The circuitry here meets all these requirements.

Once all the performance imperatives are addressed, the extra degrees of freedom can be used to, say, make components the same value for ease of procurement. Opamp circuitry is used here, apart from the hybrid

moving-coil stage. The great advantage is that all the tricky details of distortion-free amplification are confined within the small black carapace of a 5532.

One route to low noise is low-impedance design. By minimising circuit resistances the contribution of Johnson noise is reduced, and hopefully conditions set for best semiconductor noise performance. This notion is not exactly new – as some manufacturers would have you believe – but has been used explicitly in audio circuitry for at least 15 years.

In the equalisation and AGS stages, gains of much less than one are sometimes required. In these cases, avoiding the evils of attenuation-then-amplification (increased noise) and amplification-then-attenuation (reduced headroom) requires the use of a shunt feedback configuration. In the classic unity-gain stage, the shunt amplifier works at a noise gain of $\times 2$, as opposed to unity, so using shunt feed-back introduces a noise compromise at a very fundamental level.

Absolute phase is preserved for all input and outputs.

## The preamp gain structure

Compared with Ref. 1, the moving-magnet disc amplifier gain has been increased from +26 to +29 dB (all levels are at 1 kHz) to bring the line-out level up to 150 mV nominal. This is done to match equipment levels that appear to have reached some sort of consensus on this value. The input buffer has a gain of +1.0 dB with balance central.

The maximum gain of the AGS is therefore reduced from +26 to +22 dB, to retain the same maximum output of 2 V. This affects only the upper part of the gain characteristic.

## Disc input

While vinyl as a music-delivery medium is almost as obsolete as wax cylinders, there remain many sizable album collections that it is impractical to either replace with CDs or transfer to digital tape. Disc inputs must therefore remain part of the designer's repertoire for the foreseeable future.

The disc stage here accepts a moving-coil cartridge input of 0.1 or 0.5 mV, or a moving-magnet input of 5 mV. It also includes a third-order subsonic filter and the capability to drive low impedances. The moving-coil stage simply provides flat gain, of either 10 or 50 times, while the moving-magnet stage performs the full RIAA equalisation for both modes.

## Moving-coil input criteria

This stage was described in detail in Ref. 2. The prime requirement is a good noise figure from a very low source impedance – here 3.3 Ω to comply with, for example, the Ortofon *MC10* cartridge. The circuit features

- triple low-$r_b$ input transistors
- two separate d.c. feedback loops
- combined feedback-network and output-attenuator.

The very low value of $R_6$ means that a series capacitor to reduce the gain to unity at d.c. is impracticable; there is no d.c. feedback through $R_7$, $R_{10}$ around the global loop. Local d.c. negative feedback via $R_2$, $R_3$ sets input transistor conditions, and dc servo $IC_2$ applies whatever is needed to $IC_1$ non-inverting input to bring *IC*1 output to 0 V.

The two gains provided are 10× and 50×, so inputs of 0.5 mV and 0.1 mV will give 5 m V r.m.s. out. The equivalent input noise of the moving-coil stage alone is −141 dBu, with no RIAA. Johnson noise from a 3.3 Ω resistor is −147 dBu, so the noise figure is a rather good 6 dB. Resistor $R_6$ is also 3.3Ω. This component generates the same amount of noise as the source impedance, which only degrades the noise figure by 1.4 dB, rather than 3 dB, as transistor noise is significant.

If discrete transistors seem like too much trouble, remember a 5532 stage here would be at least 15 dB noisier.

## The moving-magnet input stage

The first half of Morgan Jones's excellent preamp article[5] appeared just after this preamp design was finalised. While I thoroughly endorse most of his conclusions on RIAA equalisation, we part company on two points. Firstly, I am sure that 'all-in-one-go' RIAA equalisation as in Figure 2(a) is definitely the best method, for IC op-amp designs at least. In my design the resultant loss of high-frequency headroom is only 0.5 dB at 20 kHz, which I think I can live with.

Secondly, I do not accept that the difficulties of driving feedback networks with low-impedance at h.f. are insoluble. I quite agree that 'very few preamps of any age' meet a +28 dB ref 5 mV overload margin, but some exceptions are Ref. 1 with +36 dB, Ref. 3 with +39 dB, and Ref. 4 with a tour-de-force +47 dB. My design here gives +36 dB across most of the audio band, falling to +33 dB at 20 kHz (due to h.f. pole-correction) and +31 dB at 10 Hz (due to the IEC rolloff being done in the second stage).

Many contemporary disc inputs use an architecture that separates the high and low RIAA sections. Typically there is a low-frequency RIAA stage

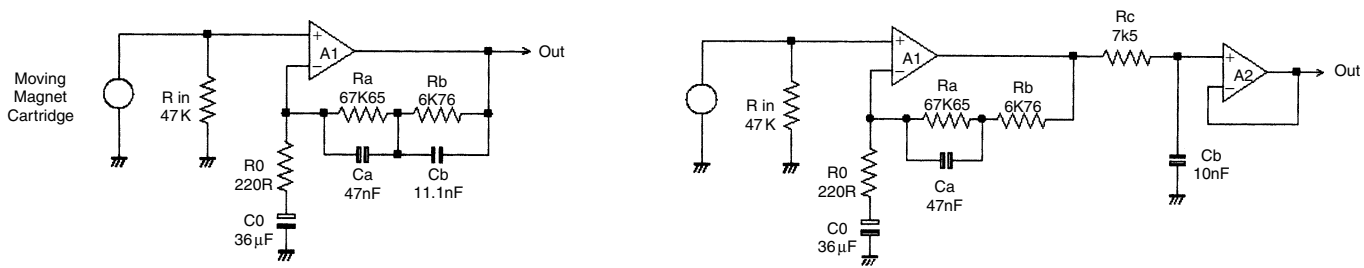



**Figure 2** *The basic RIAA configurations. Figure 2(a) is the standard 'all-in-one-go' series feedback configuration; the values shown do not give accurate RIAA equalisation. Figure 2(b) is the most common type of passive RIAA, with a headroom penalty of* 14 dB *at* 10 kHz.

followed by a passive h.f. cut beginning at 2 kHz, Figure 2(b). The values shown give a correct RIAA curve.

Amplification followed by attenuation always implies a headroom bottleneck, and passive h.f. cut is no exception. Signals direct from disc have their highest amplitudes at high frequencies so this passive configuration gives poor h.f. headroom. Overload occurs at A1 output before passive h.f. cut can reduce the level.

Figure 3 shows how the level at A1 output (Trace B) is higher at h.f. than the output signal (Trace A). Trace C shows the difference, i.e. the headroom loss; from 1 dB at 1 kHz this rises to 14 dB at 10 kHz and continues to increase in the ultrasonic region. The passive circuit was driven from an inverse RIAA network. Using this, a totally accurate disc stage would give a straight line just below the +30 dB mark.

A related problem is that A1 in the passive version must handle a signal with much more h.f. content than A1 in Figure 2(a). This worsens any difficulties with slew-limiting and h.f. distortion. The passive version uses two amplifier stages rather than one, and more precision components.

Another difficulty is that A1 is more likely to run out of open-loop gain at h.f. This is because the response plateaus above 1 kHz, rather than being

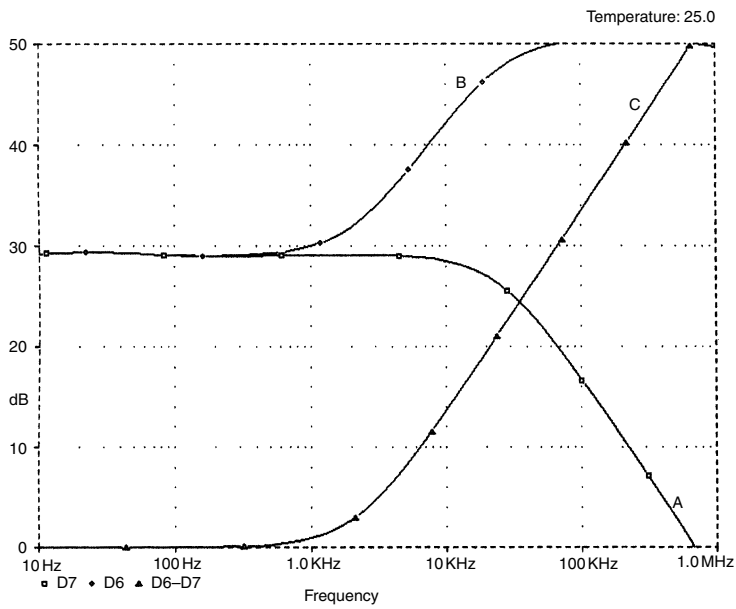



**Figure 3** *Headroom loss with passive RIAA equalisation. The signal at A1 (Trace B) is greater than A2 (Trace A) so overload occurs there. The headroom loss is plotted as Trace C.*

steadily reduced by increasing negative feedback. Passive RIAA is not an attractive option.

Alternatively there may be a flat input stage followed by a passive h.f. cut and then another stage to give the I.f. boost, which has even more headroom problems and uses yet more bits. The 'all-in-one-go' series feedback configuration in Figure 2(a) avoids unnecessary headroom restrictions and has the minimum number of stages.

## In search of accurate RIAA

I have a deep suspicion that such popularity as passive RIAA has is due to the design being much easier. The time-constants are separate and non-interactive; only the simplest of calculations are required.

In contrast the series-feedback system in Figure 2(a) has serious interactions between its time-constants and design by calculation is complex. The values shown in Figure 2(a) are what you get if you ignore the interactions and simply implement the time-constants as $R_a \times C_a$ equals 3180 µs, $R_b \times C_a$ equals 318 µs, and $R_b \times C_b$ equals 75 µs. The resulting errors are ±0.5 dB ref 1 kHz.

Empirical approaches (cut-and-try) are effective if great accuracy is not required, but attempting to reach even ±0.2 dB by this route becomes very tedious and frustrating. Hence the Lipshitz equations[6] have been converted to a spreadsheet, and used to synthesise the design in Figure 8.

A great deal of rubbish has been talked about RIAA equalisation and transient response, in perverse attempts to render the shunt RIAA configuration acceptable despite its crippling 14 dB noise disadvantage. The heart of the matter is that the RIAA replay characteristic apparently requires the h.f. gain to fall at a steady 6 dB/octave forever.

A series-feedback disc stage-with relatively low gain cannot make its gain fall below one, and so the 6 dB/octave fall tends to level out at unity early enough to cause errors in the audio band. Adding a high-frequency correction pole – i.e. low-pass time constant – just after the input stage makes the simulated and measured frequency response identical to a shunt-feedback version, and retains the noise advantage.

At this level of accuracy, the finite gain open-loop gain of even a 5534 at h.f. begins to be important, and the frequency of the h.f. pole is trimmed to allow for this.

What RIAA accuracy is possible without spending a fortune on precision parts? The best tolerance readily available for resistors and capacitors is ±1%, so at first it appears that anything better than ±0.1 dB accuracy is impossible. Not so. The component-sensitivity plots in Figures 4, 5 show the effect of 1% deviations in the value of $R_a$, $R_b$; the response errors never

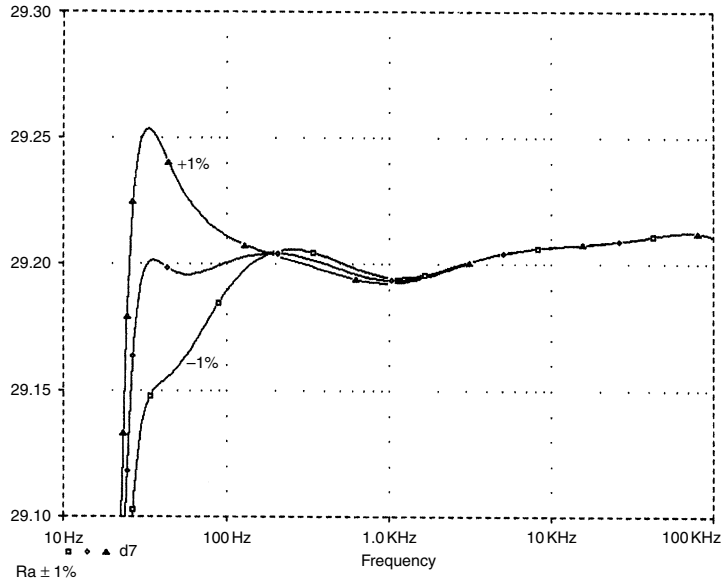



**Figure 4** *The effect on RIAA accuracy of a* ±1% *variation in* $R_a$. *Worst-case is* 0.05 dB, *only significant below* 100 Hz.

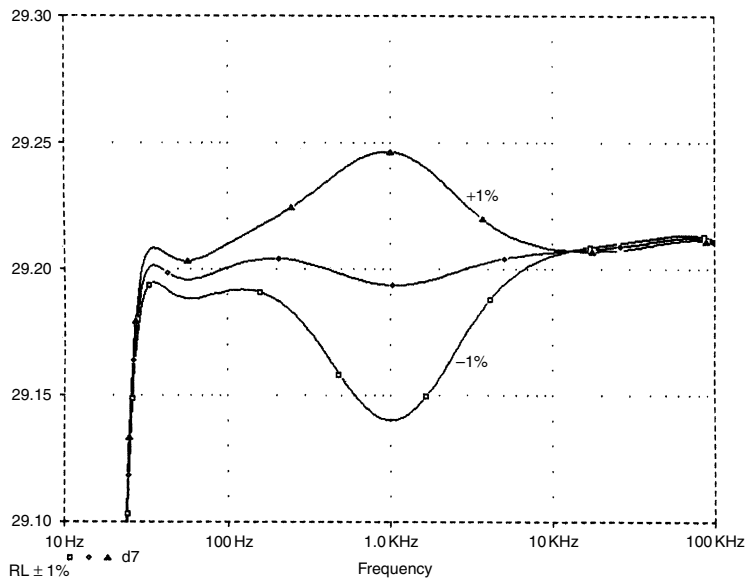



**Figure 5** *The effect on RIAA accuracy of a* ±1% *variation in* $R_b$. *Worst-case* 0.05 dB *around* 1 kHz.

exceed 0.05 dB, as there are always at least two components contributing to the RIAA response.

Sensitivity of the RIAA capacitors is shown in Figures 6, 7 and you can see that tighter tolerances are needed for $C_a$ and $C_b$, than for $R_a$ and $R_b$ to produce the same 0.05 dB accuracy. The capacitors have more effect on the response than the resistors.

Finding affordable close-tolerance capacitors is not easy; the best solution seems to be, as in 1983, axial polystyrene, available at 1% tolerance. These only go up to 10 nF, so some parallelling is required, and indeed turns out to be highly desirable. The resistors are all 1%, which is no longer expensive or exotic, though anything more accurate certainly would be.

For $C_a$, the five 10 nF capacitors in parallel reduce the tolerance of the combination to 0.44%. This statistical trick works because the variance of equal summed components is the sum of the individual variances. Thus for five 10 nF capacitors, the standard deviation (square root of variance) increases only by the square root of five, while total capacitance has increased five times. This produces an otherwise unobtainable 0.44% close-tolerance 50 nF capacitor.

Similarly, $C_b$ is mainly composed of three 4n7 components and its tolerance is improved by root-three, to 0.58%.

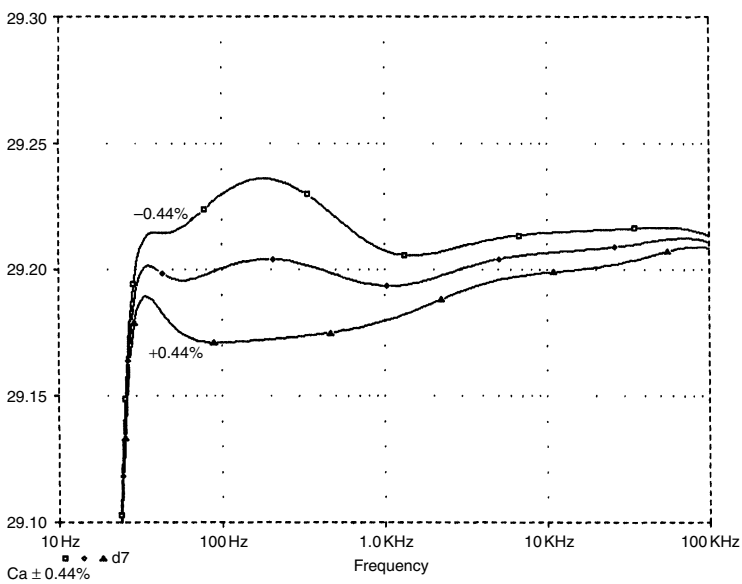



**Figure 6** *The effect on RIAA accuracy of a* ± 0.44% *variation in* $C_a$. *Effect is less than* ± 0.05 dB *at low frequencies, with a small effect on the upper audio band.*

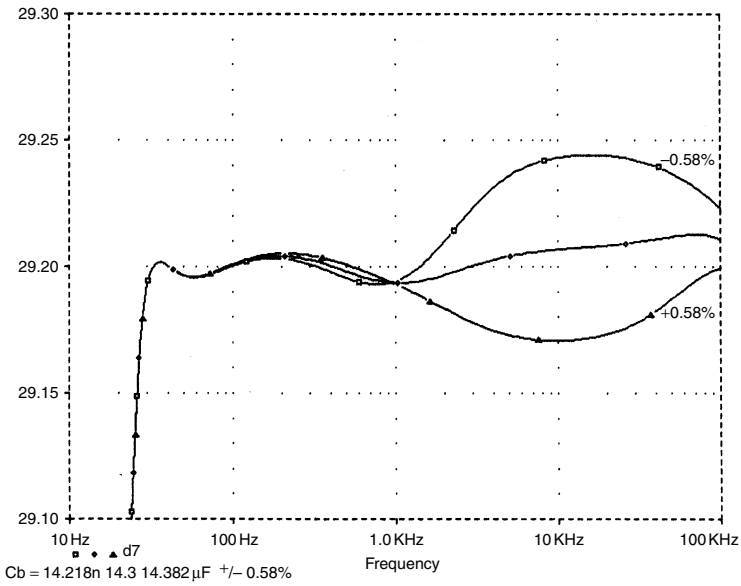



**Figure 7** *The effect on RIAA accuracy of a* ±0.58% *variation in* $C_b$. *Effect is less than* ±0.05 dB *on top four octaves. Smaller variation is permissible in the capacitors for the same RIAA error.*

## Noise considerations

The noise performance of any input stage is ultimately limited by Johnson noise from the input source resistance. The best possible equivalent input noise data for resistive sources, for example microphones with a 200 Ω source resistance, i.e. −129.6 dBu, is well-known, but the same figures for moving-magnet inputs are not.

It is particularly difficult to calculate equivalent input noise for moving magnet stages as a highly inductive source is combined with the complications of RIAA equalisation.[7] The amount by which a real amplifier falls short of the theoretical minimum equivalent input noise is the noise figure, NF. I often wonder why noise figures are used so little in audio; perhaps they are a bit too revealing.

The noise performance of disc input stages depends on the input source impedance, the cartridge inductance having the greatest influence. It is vital to realise that no value of resistive input loading will give realistic noise measurements.

A 1 kΩ load models the resistive part of the cartridge impedance. But it ignores the fact that the 'noiseless' inductive reactance makes the impedance seen at the preamp input rise very strongly with frequency, so that at higher frequencies most of the input noise actually comes from

**Table 1** *Measured noise results, showing the 5532's superiority*

| $Z_{source}$ | *TL072* | *5532* | *5532* | *5532* |
|---|---|---|---|---|
| | | | benefit | EIN |
| 1 k | −88.0 | −97.2 dBu | +9.8 dB | −126.7 dBu |
| Shure M75ED | −87.2 | −92.3 dBu | +5.1 dB | −121.8 dBu |
| (Preamp gain +29.55 dB at 1 kHz. Bandwidth 400–22 kHz, r.m.s. sensing) | | | | |

the 47 kΩ loading resistance. I am grateful to Marcel van de Gevel[8] for drawing my attention to this point.

Hence, for the lowest noise you must design for a higher impedance than you might think, and it is fortunate that the RIAA provides a treble roll-off, or the noise problem would be even worse than it is. This is not why it was introduced. The real reason for pre-emphasis/de-emphasis was to discriminate against record surface noise. Table 1 shows the two most common audio op-amps, the 5532 being definitely the best and quieter by 5 dB.

To calculate appropriate EINs, I built a spreadsheet mathematical model of the cartridge input, called MAGNOISE. The basic method is as in Ref. 9. The audio band 50–22 kHz is divided into nine octaves, allowing RIAA equalisation to be applied, and the equivalent generators of voltage noise ($e_n$) and current noise ($i_n$) to be varied with frequency.

Noise generated by the 47 kΩ resistor $R_{in}$ is modelled separately from its loading effects so its effect can be clearly seen. I switched off the bottom three octaves to make the results comparable with real cartridge

## Filtering subsonics

This stage is a third-order Butterworth high-pass filter, modified for a slow initial rolloff that implements the IEC amendment. This is done by reducing the value of $R_{27} + R_{28}$ below that for maximal flatness. The stage also buffers the high-frequency correction pole, and gives the capability to drive a 600 Ω load, if you can find one.

Capacitor distortion[10] in electrolytics is – or should be – by now a well-known phenomenon. It is perhaps less well known that non-electrolytics can also generate distortion in filters like these. This has nothing to do with Subjectivist musicality, but is very real and measurable.

The only answer appears to be using the highest-voltage capacitors possible; 100 V polyester generates ten times less distortion than the 63 V version.

measurements that require a 400 Hz high-pass filter to eliminate hum, and 1/f effects are therefore neglected. No psychoacoustic weighting was used, and cartridge parameters were set to 610 Ω + 470 mH, the measured values for the Shure *M75ED*.

The results match well with my 5532 and *TL072* measurements, and I think the model is a usable tool. Table 2 shows some interesting cases; output noise is calculated for gain of +29.55 dB at 1 kHz, and signal-to-noise ratio for a 5 mV r.m.s. input at 1 kHz.

I draw the following conclusions. The minimum equivalent input noise from this particular cartridge, without the extra thermal noise from the 47 kΩ input loading, is −133.5 dBu, no less than 7 dB quieter than the loaded cartridge. (Case 1) It is the quietest possible condition. The noise difference between 10 MΩ and 1 MΩ loading is still 0.2 dB, but as loading resistance is increased further to 1000 MΩ the EIN asymptotes to −133.5 dBu. A 47 kΩ loading is essential for correct cartridge response.

With 47 kΩ load, the minimum EIN from this cartridge is −126.5 dBu. (Case 5) All other noise sources, including $R_0$, are ignored. This is the appropriate noise reference for this preamp design.

Resistor $R_0$, the 220 Ω resistor in the bottom arm of negative feedback network, adds little noise. The difference between Case 5 and Case 7 is only 0.3 dB.

A disc preamp stage using a good discrete bipolar device such as the remarkable *2SB737* transistor ($r_b$ only 2 Ω typ) is potentially 2.8 dB quieter than a *5532*, when the noise from $R_0$ and the input load are included. Compare Cases 11 and 16.

**Table 2** *Calculated minimum noise results*

| *Case* | $e_n$ | $i_n$ | $R_{in}$ | $R_o$ (R) | *Output* | *S/N ref* (dB) | *EIN* |
|---|---|---|---|---|---|---|---|
| | nV/√Hz | pA/√Hz | | | dBu | 5 mV | dBu |
| 1 Noiseless amp | 0 | 0 | 1000 *M* | 0 | −104.0 | −89.7 | −133.5 A |
| 5 Noiseless amp | 0 | 0 | 47 k | 0 | −97.1 | −82.8 | −126.5 C |
| 7 Noiseless amp | 0 | 0 | 47 k | 220 | −96.7 | −82.4 | −126.2 |
| 11 *2SB737*, $I_c = 70$ μA | 1.7 | 0.4 | 47 k | 220 | −95.3 | −81.0 | −124.8 |
| 16 5532 | 5 | 0.7 | 47 k | 220 | −92.5 | −78.2 | −122.0 |
| 18 TL072 | 18 | 0.01 | 47 k | 220 | −86.9 | −72.6 | −116.5 |

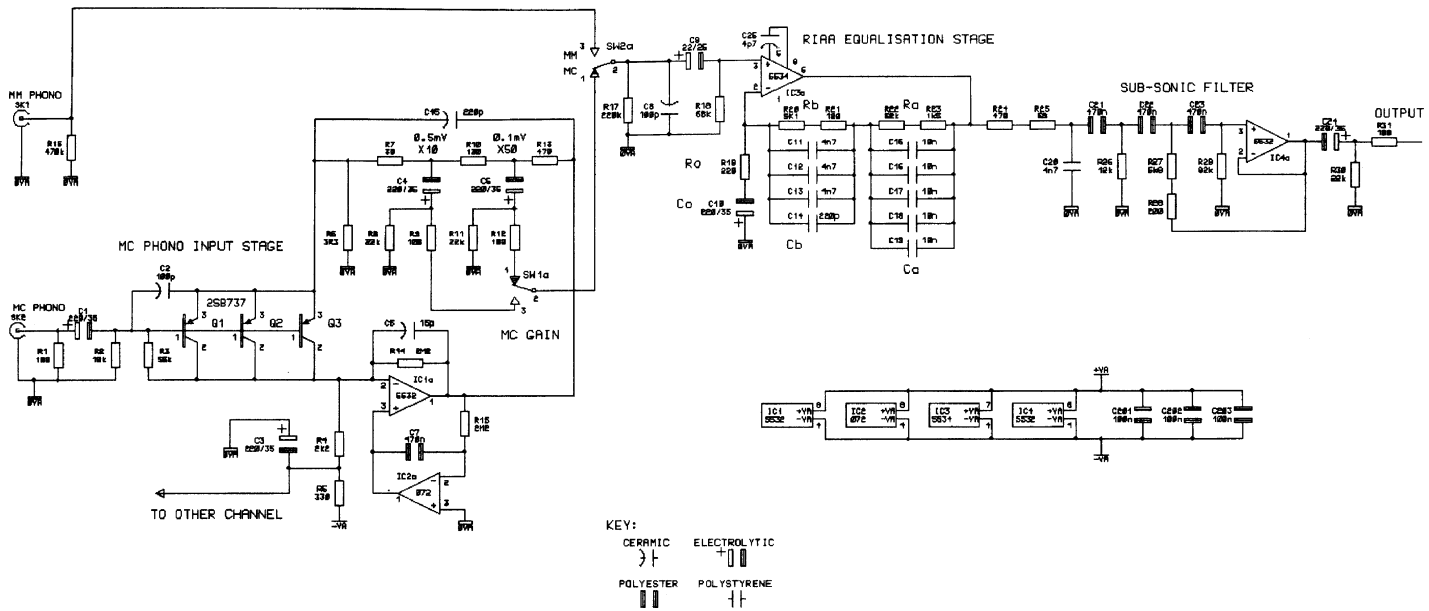



**Figure 8** *Switchable for moving coil or moving-magnet type cartridges, the disc amplifier includes a subsonic filter to reduce cone excursions and distortion due to warped vinyl.*

The calculated noise figure for a 5532 is 4.5 dB. Measured noise output of the moving magnet stage is −92.3 dBu (1 kHz gain +29.5 dB) and so the equivalent input noise is −121.8 dBu, and the real noise figure is 4.7 dB, which is not too bad. Noise from the subsonic filter is negligible.

Taking $e_n$ and $i_n$ from data books, it looks as though the 5534/5532 is the best op-amp possible for this job. Other types – such as *OP-27* – give slightly lower calculated noise, but measure slightly higher. This is probably due to extra noise generated by bias current-cancellation circuitry.[8]

There is an odd number of half-5532s, so the single 5534 is placed in the moving-magnet stage, where its slightly lower noise is best used. The RIAA-equalised noise output from the disc stage in moving-coil mode is −93.9 dBu for 10× times gain, and −85.8 dBu for 50× times. In the 10× case the moving-coil noise is actually 1.7 dB lower than moving-magnet mode.

## Circuit details

The complete circuit of the disc amplifier and subsonic filter is Figure 8. Circuit operation is largely described above, but a few practical details are added here. Resistors $R_9$ and $R_{12}$ ensure stability of the moving-coil stage when faced with moving-magnet input capacitance $C_8$, while $R_8$ and $R_{11}$ are dc drains.

The 5534 moving-magnet stage has a minimum gain of about 3×, so compensation should not be required; if it is, a position is provided ($C_{26}$) for external capacitance to be added; 4.7 pF should be ample. The moving magnet stage feedback arm $R_{20-23}$ has almost exactly the same d.c. resistance as the input bias resistor $R_{18}$, minimising the offset at the output of $IC_3$. The h.f. correction pole is $R_{24} + R_{25}$ and $C_{20}$.

Capacitor $C_{24}$ is deliberately oversized so low loads can be driven. Resistor $R_{31}$ ensures stability into high-capacitance cables.

## References

1. Self, D. 'A precision preamplifier', *Wireless World*, October 1983, p 31.
2. Self, D. 'Design of moving-coil head amplifiers', *Electronics & Wireless World*, December 1987, p 1206.
3. Self, D. 'An advanced preamplifier design', *Wireless World*, November 1976, p 41.
4. Self, D. 'High-performance preamplifier', *Wireless World*, February 1979, p 43.

5. Jones, M. 'Designing valve preamps', *Electronics World*, March 1996, p 192.
6. Lipshitz, S. 'On RIAA equalisation networks', *Journ Audio Eng Soc*, June 1979, p 458.
7. Walker, H.P. 'Low noise amplifiers', *Wireless World*, May 1972, p 233.
8. van de Gevel, M. Private communication, February 1996.
9. Sherwin, J. 'Noise specs confusing?', National Semiconductor Application Note AN-104, Linear Applications Handbook 1991.
10. Self, D. 'Capacitor distortion, Views', *HiFi News & RR*, November 1985, p 23.

# 6 Precision preamplifier '96, Part II

*September 1996*

Morgan Jones raised the excellent point of crosstalk in the input-select switching in a recent article (Valve Preamplifiers, *Electronics World*, March/April 1996). If the source impedance is significant then this may be a serious problem.

While I agree that Morgan's rotary switch with every other contact grounded may be slightly superior to conventional rotary switches, measuring a popular Lorlin switch type showed the improvement to be only 5 dB. I am also unhappy with all those redundant 'mute' positions between input selections, so I instead chose interlocked push-switches rather than a rotary. A four-pole-changeover format can then be used to reduce crosstalk.

The problem with conventional input select systems like Figure 1(a) is that the various input tracks necessarily come into close proximity, with significant crosstalk through capacitance $C_{stray}$ to the common side of the switch, i.e. from A to B. Using two changeovers per input side – i.e. four for stereo – allows the intermediate connection B-C to be grounded by the NC contact of the first switch section. This keeps the 'hot' input A much further away from the common input line D, as shown in Figure 1(b).

Crosstalk data in Table 1 was gathered at 10 kHz, with 10 kΩ source impedances. The emphasis here is on minimising inter-source crosstalk, as interchannel (L–R) crosstalk is benign by comparison. Interchannel isolation is limited by the placement of left and right on the same switch, with the contact rows parallel, and limits L–R isolation to −66 dB at 10 kHz with 10 kΩ source impedance.

With lower source impedances, both intersource and interchannel crosstalk is proportionally reduced. In this case, a more probable 1 kΩ source gives 115 dB of intersource rejection at 10 kHz for the four-pole-changeover method.

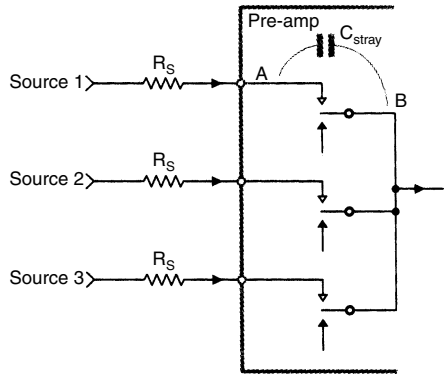



**Figure 1(a)** *Input-select switching for audio preamplifiers – the conventional method, with poor rejection of unselected sources due to* $C_{stray}$.

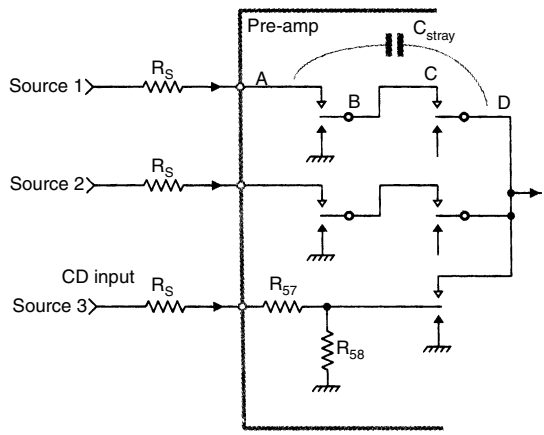



**Figure 1(b)** *Improved input selection using four-pole switching to reduce capacitance between the different sources. The CD input attenuator can be grounded when not selected, so two-pole switching is sufficient for high isolation of Source 3.*

**Table 1** *Crosstalk exhibited by four switch arrangements using a 10 kHz test signal*

| | |
|---|---|
| Simple rotary: | 71 dB |
| Morgan-Jones rotary: | 76 dB |
| 2 c/o switch | 74 dB |
| 4 c/o switch | 95 dB |

## Line input criteria

Nowadays the input impedance of a preamp must be high to allow for interfacing to anachronistic valve equipment, whose output may be taken from a valve anode. Even light loading compromises distortion and available output swing. A minimum input impedance is 100 kΩ, which many preamp designs fall well short of.

The CD input stands out from other line sources in that its nominal level is usually 1 V rather than 150 mV. This is perfectly reasonable, since digital sources have rigidly defined maximum output levels, and these might as well be high to reduce noise troubles. There is no danger of the analogue output section clipping. However, this means a direct line input cannot be used without the trouble of resetting volume and recording-level controls whenever the CD source is selected.

This problem is addressed here by adding a 16 dB passive attenuator, as shown for Source 3 in Figure 1(b). The assumption is that a CD output has a low impedance, and that a 10 kΩ input impedance will not embarrass it. As a result, resistance values can be kept low to minimise the noise degradation. Output impedance of this attenuator is 1.4 kΩ, which generates −120.9 dBu of Johnson noise as opposed to −135.2 dBu from a direct 50 Ω source. This is still much less than the preamp internal noise and so the noise floor is not degraded. It is now possible to improve inter-source crosstalk simply by grounding the CD attenuator output when it is not in use, so only a two-pole switch is required for good isolation of this source.

The tape-monitor switch allows the replay signal from the tape deck to be compared with the source signal. With three-head machines, this provides a real-time quality check. But with the much more common two-head appliances, where the input signal is looped straight back to the amplifier in RECORD mode, it only provides confirmation that the signal has actually got there and back.

## Line input buffering

This stage has to provide a high input impedance and variable gain for the balance control. My last preamp[1] had the balance control incorporated in the tone stage, but this does not appear to be practical with the more complex tone system here.

The vernier balance control alters the relative stage gain by +4.5, −1.1 dB – a difference of 6 dB – which is sufficient to swing the image wholly from one side to the other. Since the minimum gain of this non-inverting stage is unity, the nominal gain with balance control central is 1.1 dB. Maximum gain of the active gain stage, or AGS, is reduced to allow

for this. The active nature of this balance control means that the signal never receives unwanted attenuation that must be undone later with noisy amplification. The gain law is modified by $R_{34}$ to give as little gain as possible in the centre. Maximum gain is set by $R_{35}$, Figure 11(b).

A high input impedance is obtained simply by using a high-value biasing resistor $R_{33}$, accepting that the bias current through this will give some negative output offset; at −180 mV this is not large enough to reduce headroom. Input impedance is therefore 470 kΩ, high enough to prevent loading problems with any conceivable source equipment.

In discussing noise there are fundamental limits that lend perspective to the process. If the external source impedance is 50 Ω, which is about as low as is plausible, the inherent thermal noise from it is −135.2 dBu in 22 kHz bandwidth. This is well below the measuring equipment, (AP System 1) which has an input noise floor I measured at −116.8 dBu, 50 Ω source again.

The noise output of the buffer/balance stage is of the same order and cannot be measured directly – a good way would be to use the flat moving-coil cartridge stage as a preamp for the testgear.[2] Calculated noise output is −116 dBu with balance central.

## Controlling tone

I plan to ignore convention once again. I think tone controls are absolutely necessary, and it is a startling situation when, as frequently happens, anxious inquirers to hi-fi advice columns are advised to change their loudspeakers to correct excess or lack of bass or treble. This is an extremely expensive way of avoiding tone controls.

This design is not a conventional Baxandall tone control. The break frequencies are variable over a ten to one range, because this makes the facility infinitely more useful for correcting speaker deficiencies. This enhancement flies in the face of Subjectivist thinking, but I can live with that. Variable boost/cut and frequency enables any error at top or bottom end to be corrected to at least a first approximation. It makes a major difference, as anyone who has used a mixing console with comprehensive EQ will tell you.

Middle controls are quite useless on a preamplifier. They are no good for acoustic correction: after all, even a third-octave graphic equaliser isn't that much use. Variable frequency mid controls are standard on mixing consoles because their function is voicing – i.e. giving a sound a particular character – rather than correcting response anomalies.

Certain features of the tone control may make it more acceptable to those with doubts about its sonic correctness. The tone control range is restricted to ±10 dB, rather than the ±15 dB which is standard in mixing

consoles. The response is built entirely from simple 6 dB/octave circuitry, with inherently gentle slopes. The stage is naturally minimum-phase, and so the amplitude curves uniquely define the phase response. This will be shown later, where the maximum phase-shift does not exceed 40° at full boost.

This is a return-to-flat tone control. Its curves do not plateau or shelve at their boosted or cut level, but smoothly return to unity gain outside the audio band. Boosting 10 kHz is one thing, but boosting 200 kHz is quite another, and can lead to some interesting stability problems. The fixed return-to-flat time-constants mean that the boost/cut range is necessarily less at the frequency extremes, where the effect of return-to-flat begins to overlap the variable boost/cut frequencies.

The basic principle is shown in Figure 2. The stage gives a unity-gain inversion, except when the selective response of the side-chain paths allow signal through. In the treble and bass frequency ranges, where the side-chain does pass signal, boost/cut potentiometers $VR_{2,4}$ can give either gain or attenuation. When a wiper is central, there is a null at the middle of the boost/cut potentiometer, no signal through that side-chain, and gain is unity.

If the potentiometer is set so the side-chain is fed from the input then there is a partial cancellation of the forward signal; if the side-chain is fed

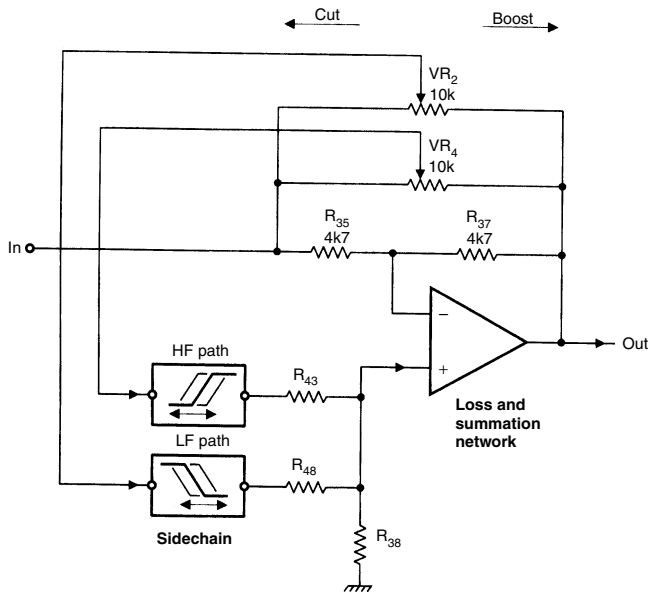



**Figure 2** *The basic tone control circuit. The response only deviates from unity gain at frequencies passed by the h.f. or if side-chain paths.*

from the output then there is a partial negative-feedback cancellation. To put it another way, positive feedback is introduced to counteract part of the negative feedback through $R_{37}$.

This apparently ramshackle process actually gives boost/cut curves of perfect symmetry. In fact this symmetry is pure cosmetics, because you can't use both sides of the curve at once, so it hardly matters if they are exact mirror-images.

## Bass and treble

The tone control stage acts in separate bands for bass and treble, so there are two parallel selective paths in the side-chain. These are simple *RC* time-constants, the bass path being a variable-frequency first-order low-pass filter, and the associated bass control only acting on the frequencies this lets through.

Similarly, the treble path is a variable high-pass filter. The filtered signals are summed and returned to the main path via the non-inverting input, and some attenuation must be introduced to limit cut and boost.

Assuming a unity-gain side-chain, this loss is 9 dB if cut and boost are to be limited to ±10 dB. This is implemented by $R_{43}$, $R_{48}$ and $R_{38}$, Figures 2, 3 and 4. The side-chain is unity-gain, and so has no problems with clipping before the main path does. As a result, it is highly desirable to put the loss after the sidechain, where it attenuates side-chain noise.

The loss attenuator is made up of the lowest value resistors that can be driven without distortion. This minimises both the Johnson noise therein and noise generated by op-amp $IC_{7b}$.

The tone cancel switch disconnects the entire sidechain, i.e. five out of six op-amps, from any contribution to the main path, and usefully reduces

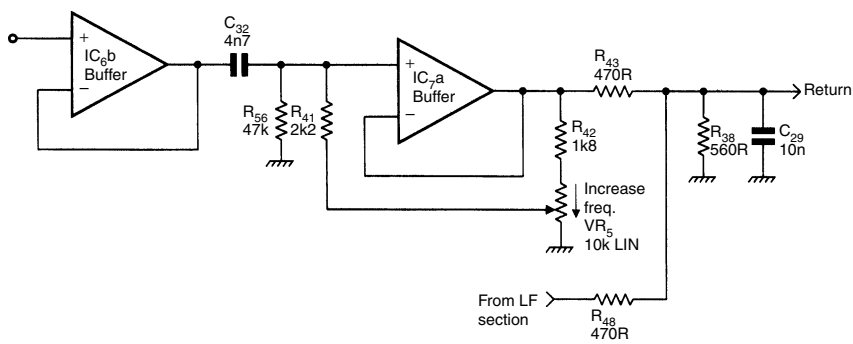



**Figure 3** *The treble frequency control circuit, with a range of* 1 *to* 10 kHz. *The variable bootstrapping of* $R_{41}$ *via* $VR_5$ *renders the control law approximately logarithmic.*

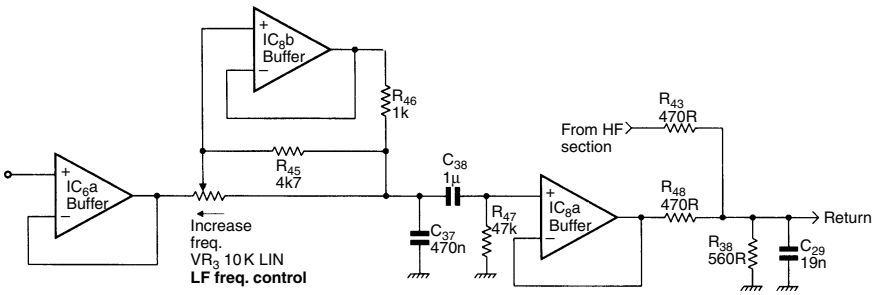



**Figure 4** *The bass frequency control circuit for* 100 Hz *to* 1 kHz. $R_{46}$ *aids* $VR_3$ *in driving* $C_{37}$ *by an amount that varies with the control setting.*

the stage output noise by about 4 dB, depending on the h.f. frequency setting. It leaves only $IC_{7b}$ in circuit, which is required anyway to undo the gain-control phase-inversion.

Unlike configurations where the entire stage is by-passed, the signal does not briefly disappear as the switch moves between two contacts. This minimises transients due to suddenly chopping the waveform and makes valid tone in/out comparisons much easier.

Having all potentiometers identical is very convenient. I have used linear 10 kΩ controls, so the tolerances inherent in a two-slope approximation to a logarithmic law can be eliminated. This only presents problems in the tone stage frequency controls, as linear potentiometers require thoughtful circuit design to give the logarithmic action that fits our perceptual processes.

Basics of the treble path are shown in Figure 3. Components $C_{32}$, $R_{41}$ are the high-pass time-constant, driven at low-impedance by unity-gain buffer $IC_{6b}$. This is needed to prevent the frequency from altering with the boost/cut setting. The effective value of $R_{41}$ is altered over a 10:1 range by varying the amount of boot-strapping it receives-from $IC_{7a}$, the potential divider effect and the rise in source resistance of $VR_5$ in the centre combining to give a reasonable approximation to a logarithmic frequency/rotation law, Figure 5.

Resistor $R_{42}$ is the frequency end-stop resistor. It limits the maximum effective value of $R_{41}$. Capacitor $C_{29}$ is the treble return-to-flat capacitor. At frequencies above the audio band it shunts all the sidechain signal to ground, preventing the treble control from having any further effect.

The treble side-chain does degrade the noise performance of the tone control stage by 2–3 dB when connected. This is because it must be able to make a contribution at the h.f. end of the audio band. As you would expect, the noise contribution is greatest when the h.f. frequency is set to minimum, and so a wider bandwidth from the side-chain contributes to the main path.

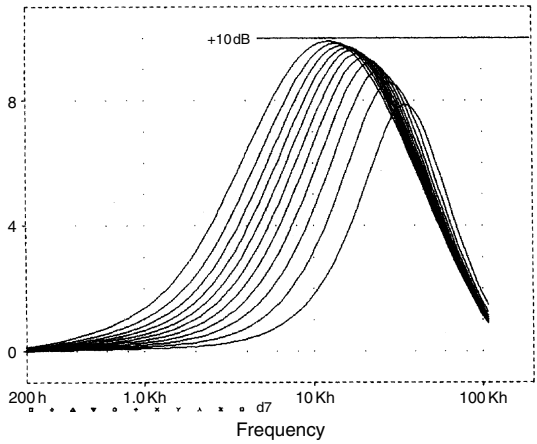



**Figure 5** *Treble frequency control law for constant increments of rotation. The curves approximate to linear spacing on the log frequency axis. PSpice simulation.*

The simplified bass path is shown in Figure 4. Op-amp $IC_{6a}$ buffers $VR_2$ to prevent boost/frequency interaction. The low-pass time-constant capacitor is $C_{37}$, and the resistance is a combination of $VR_3$ and $R_{45,46}$.

Capacitors $C_{38,39}$ with $R_{47}$ make up the return-to-flat time-constant for the bass path, which blocks very low frequencies, limiting the lower extent of bass control action. The bass frequency law is made approximately logarithmic by $IC_{8b}$; for minimum frequency $VR_3$ is set fully counter-clockwise, so the input of buffer $IC_{8b}$ is the same as the $C_{37}$ end of $R_{46}$, which is thus bootstrapped and has no effect.

## Turnover

When $VR_3$ is fully clockwise, $R_{45,46}$ are effectively in parallel with $VR_3$ and the turnover frequency is at a maximum. Resistor $R_{45}$ provides some extra law-bending, Figure 6. Sadly, an extra op-amp is required. However, despite its three op-amps, the bass side-chain contributes very little extra noise to the tone stage. This is because most of its output is inherently rolled off by the low-pass action of $C_{37}$ at high frequencies, almost eliminating its noise contribution.

Once the active elements have been chosen – here *5532*s – and the architecture made sensible in terms of avoiding attenuation-then-amplification, keeping noise-gain to a minimum, and so on, there remains one further means of improving noise performance. This is to reduce the impedance of the circuitry.

The resistances are lowered in value, with capacitances scaled up to suit, by a factor that is limited only by op-amp drive capability. This is another good reason to use the *5534/2*.

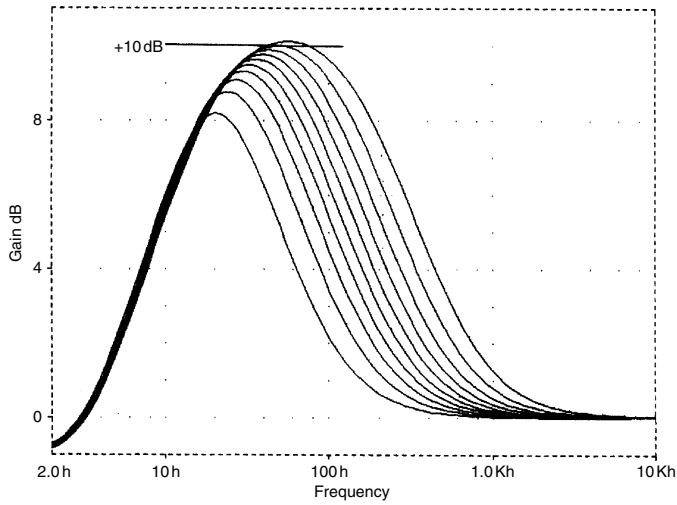



**Figure 6** *Bass frequency control law with near-linear spacing on the log frequency axis.*

Two examples of this process as applied to the tone stage are given here. In each case the noise improvement is for the stage in isolation, set flat with high frequency set at minimum:

Firstly, in this sort of stage $R_{36,37}$ are conventionally 22 kΩ. This was reduced to 4.7 kΩ, and noise output dropped by 1.3 dB. Second, the summation/loss network began with $R_{43,48}$ as 4.7 kΩ, and $R_{38}$ as 5.6 kΩ. Reducing this by a factor of ten to 470 Ω and 560 Ω respectively reduced output noise by 0.6 dB.

With balance control central and tone cancel pressed, noise output of the tone stage, plus the line/balance buffer before it is −107.2 dBu. This is 22 kHz bandwidth. With tone controls active but set flat, noise output at minimum high frequency is −104.7 dBu, and at maximum is −106.7 dBu.

The final tone stage may look rather a mess of pottage, and be afflicted with more buffers than Clapham Junction. This is unavoidable if control interaction is to be wholly eliminated. Sadly, the practical tone circuit is somewhat more complex than Figures 2, 3 and 4, reflecting one of the disadvantages of low-noise opamps. This is that bipolar input stages mean that the bias currents are non-negligible. They must not be allowed to flow through potentiometers if crackling noises are to be avoided when they are moved.

These bias currents also tend to be reflected in significant output offset voltages, as the source resistances for the two op-amp inputs are not normally the same. All gain-variable circuit stages therefore have their gain reduced to unity at d.c. This subject is detailed later.

Figure 7 shows the measured extremes of cut and boost at the frequency extremes. Figure 8 gives the phase-shift at h.f. while Figure 9 shows phase-shift at low frequencies. In both cases it is very modest.

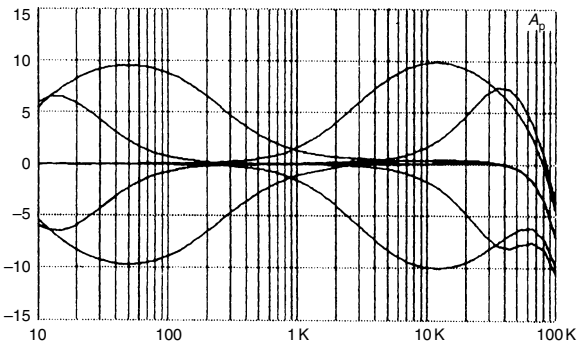



**Figure 7** *Tone-control maximum boost/cut curves (measured).*

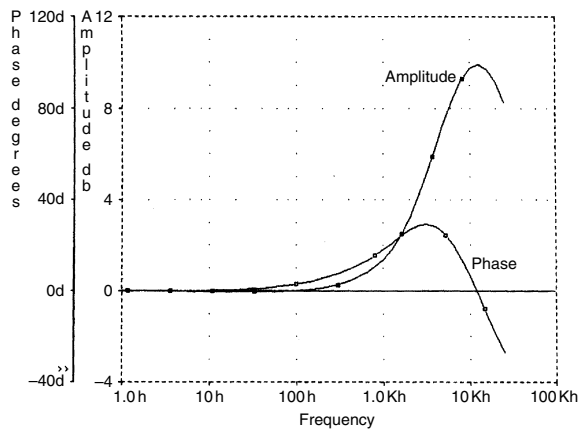



**Figure 8** *Tone control phase curve for maximum treble boost. Maximum phase-shift is* 29° *at about* 4 kHz. *PSpice simulation.*

## Active gain stage

The active gain stage, or AGS, used here as in,[1] is due to Baxandall.[3] Maximum gain is set to +23 dB by the ratio of $R_{52,53}$, to amplify a 150 mV line input to 2 V with a small safety margin.

An active volume-control stage gives the usual advantages of lower noise at gain settings below maximum, and for the Baxandall configuration, excellent channel balance that depends solely on the mechanical alignment of the dual linear potentiometer. All mismatches of its electrical characteristics are cancelled out, and there are no quasi-log dual slopes to induce anxiety.

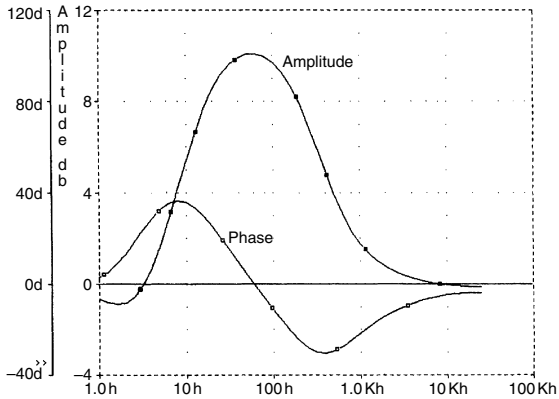



**Figure 9** *Tone-control phase curve for maximum bass boost. Maximum phase-shift is* 31° *at* 40 Hz.

Note that all the potentiometers are 10 kΩ linear types and identical, apart from the question of centre-detents, which are desirable only on the balance, treble and bass boost/cut controls.

Compared with,[1] noise has been reduced by an impedance reduction on the gain-definition network $R_{52,53}$. The limit on this is the ability of buffer $IC_{5a}$ to drive $R_{52}$, which has a virtual earth at its other end. Figure 10 shows the volume control law for different maximum gain settings; only the very top end of the curve alters significantly.

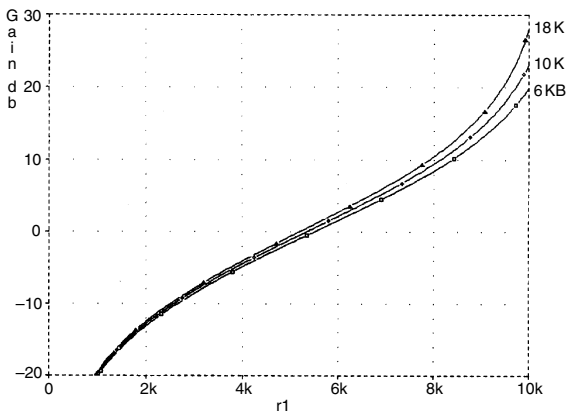



**Figure 10** *Plot of the Active Gain Stage volume control law. Varying the maximum gain has little effect except at the top end; the middle curve is the one used in the preamp.*

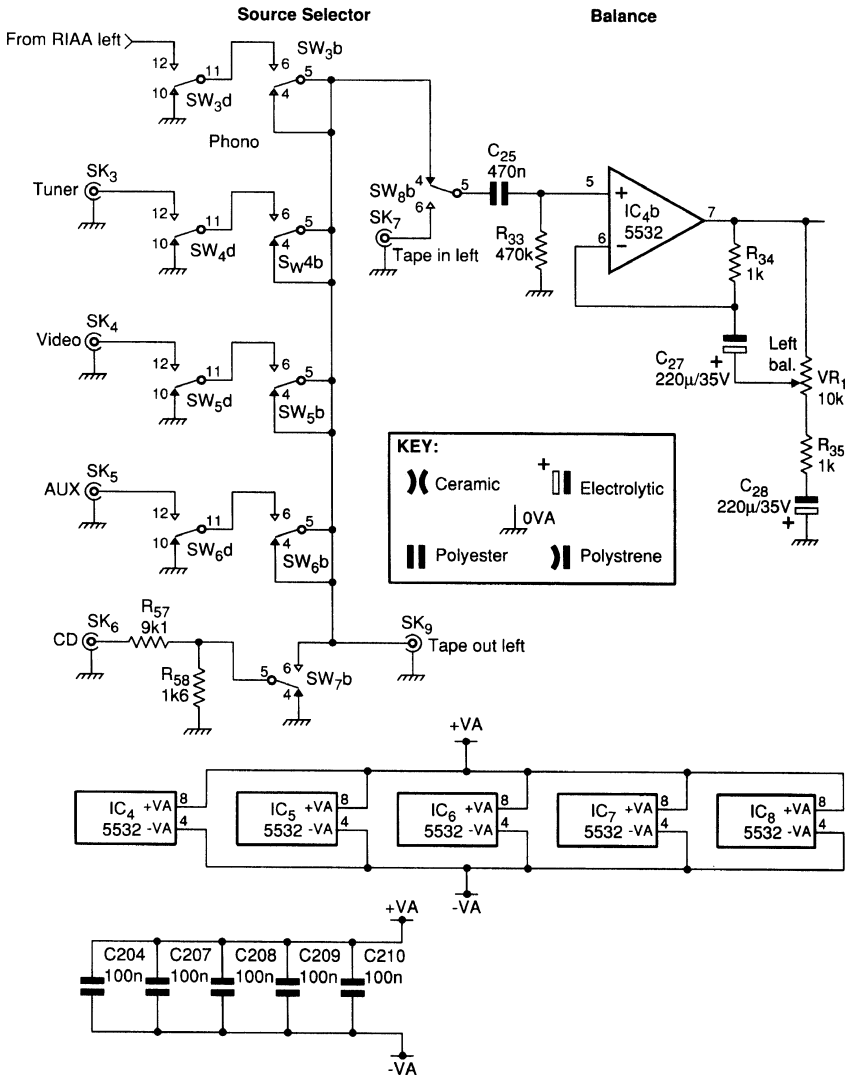



**Figure 11(a)** *Part 1 Circuit of input buffer, tone control, Active Gain Stage.*

For the rear section of the preamp – i.e. that shown in Figure 11(a) – the noise performance depends on control settings. The below gives results for h.f. frequency at minimum, the worst case, Table 2.

The figures for maximum gain may look unimpressive, but remember this is with +23 dB of gain; at normal volume settings the noise output is below −100 dBu. I think this is reasonably quiet.

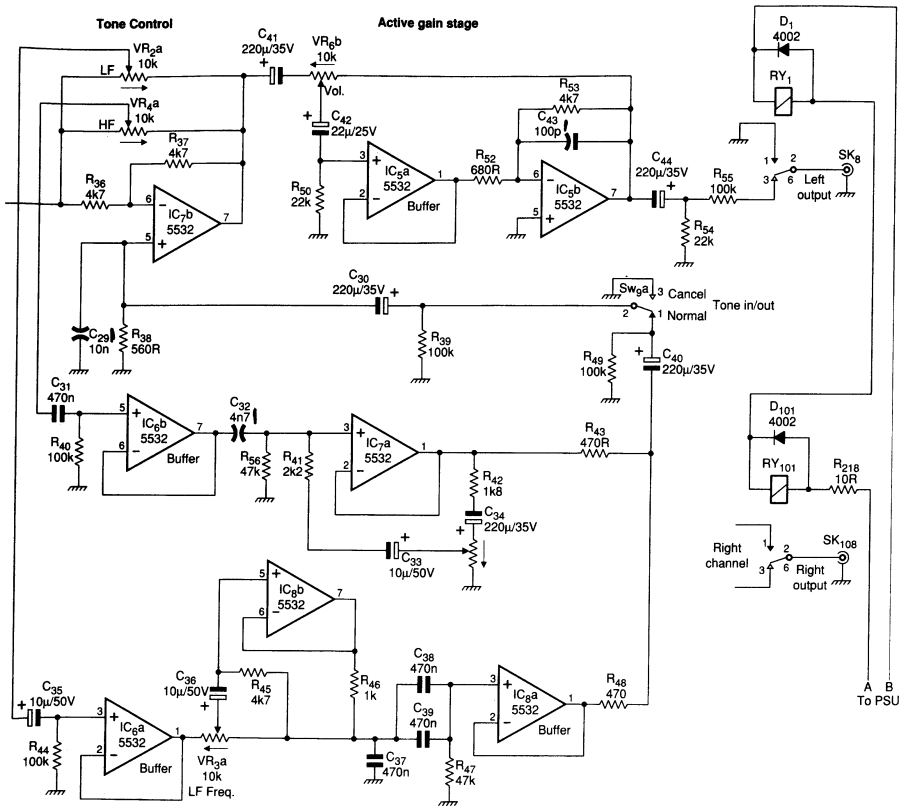



**Figure 11(a)** *Part 2 Circuit of input buffer, tone control, Active Gain Stage.*

**Table 2** *Characteristics of the tone-control stage*

| | *Tone cancel* (dBu) | *Tone flat* (dBu) |
|---|---|---|
| AGS zero gain | −114.5 | −114.5 |
| AGS unity gain | −107.4 | −105.3 |
| AGS fully up | −90.2 | −86.4 |

## Output muting and relay control

The preamp includes relay muting on the main outputs. This is to prevent thuds and bangs from upstream parts of the audio system from reaching the power amplifiers and speakers at power-up and power down. Most op-amp circuitry, being dual-rail (i.e. outputs at 0 V) does not inherently generate enormous thumps, but it cannot be guaranteed to be completely silent. It may produce a very audible turn-on thud, and often objectionable turn-off noise. I recall one design that emitted an unnerving screech of fading protest as the rails subsided . . . .

Electronic muting is desirable, but introduces unacceptable compromises in performance. Relay muting, given careful relay selection and control, is virtually foolproof. The relay must be normally-open so the output is passively muted when no power is applied. The control system must:

- Delay relay pull-in at power-up, to mute turn-on transients. A delay of at least 1 second before the relay closes.
- Drop out the relays as fast as possible at power-down, to stop the dying moans of the preamp, etc., from being audible.

My preferred technique is a 2 ms or there-abouts power-gone timer, held in reset by the a.c. on the mains transformer secondary, except for a brief period around the a.c. zero-crossing, too short to allow the timer to trigger. When the a.c. disappears, this near-continuous reset is removed, the timer fires, and relay power is removed within 2 ms. This is over long before the reservoir capacitors in the system can discharge, so turn-off transients are authoritatively suppressed.

However, if the mains switch contacts generate an r.f. burst that is in turn reproduced as a click by the preamplifier, then even this method may not be fast enough to completely mute it.

Figure 11(b) shows the practical relay-control circuit. At turn-on, $R_{211}$ slowly charges $C_{224}$ until $Tr_{205}$ and $D_{207}$ are forward biased, i.e. when $C_{224}$ voltage exceeds that set up by $R_{214,215}$. This is the turn-on delay. Transistor $Tr_{206}$ is then turned on via $R_{213}$, energising the relays, and $LD_{201}$ is brightly

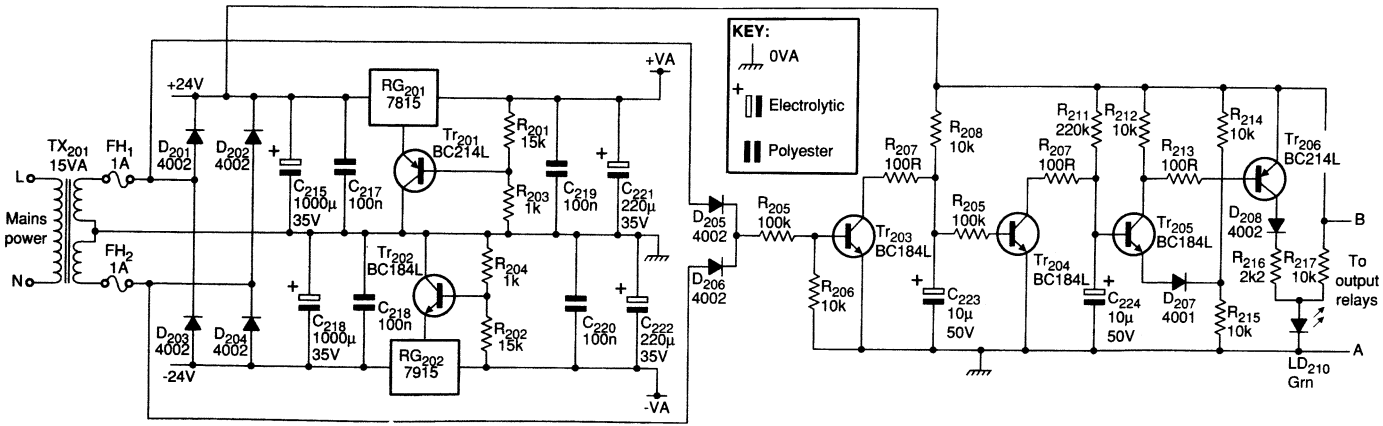



**Figure 11(b)** *Circuit of power-supply and relay controller.*

lit through $D_{208}$ and $R_{216}$. This led is dimly lit via $R_{217}$ as soon as power is applied, but only brightens when the initial mute period is over.

As long as mains power is applied, $Tr_{203}$ is kept turned on through $D_{205,206}$ by the a.c. ahead of the bridge rectifier, except during the zero-crossing period every 10 ms, when the voltage is too low for $Tr_{203}$ base to conduct. When $Tr_{203}$ switches off, $C_{223}$ starts to charge through $R_{208}$, but is quickly discharged through $R_{207}$ when the very brief zero-crossing period ends. If it does not end – in other words mains power has been switched off – $C_{223}$ keeps charging until $Tr_{204}$ turns on, discharging $C_{224}$ rapidly via $R_{210}$, and removing power from the relays almost instantly.

## DC blocking and additional details

The preamp circuitry has been described as each stage was dealt with, so this section is confined to d.c. blocking problems and other odd subjects.

The complete circuit of the line section of the preamp is Figure 11(a). Bias current is kept out of balance potentiometer $VR_1$ by $C_{27}$, and d.c. gain held to unity by $C_{28}$. Capacitors $C_{31}$ and $C_{35}$ keep bias currents out of $VR_{2,4}$, necessitating bias resistors $R_{40}$, $R_{44}$.

The treble frequency law is corrected by bootstrapping through $C_{33}$, which keeps the bias current of $IC_{7a}$ out of $VR_5$. Similarly, $C_{34}$ prevents any offset on $IC_{7a}$ output reaching $VR_5$. In the bass path $C_{36}$ keeps $IC_{8b}$ bias out of $VR_3$, while return-to-flat components $C_{38,39}$ and $R_{47}$ provide inherent d.c.-blocking.

Final offsets at the side-chain output are blocked by $C_{40}$, while $IC_{7b}$ bias is blocked by $C_{30}$. This is essential to prevent the tone-cancel switch clicking due to d.c. potentials. Bear in mind that this switch may still appear to click if it switches in or out a large amount of response-modification of a non-zero signal. This is because the abrupt gain-change generates a step in the waveform that is heard as a click. This is unavoidable with hard audio switching.

Capacitor $C_{41}$ keeps $IC_{7b}$ output offset from volume control $VR_6$, while $C_{42}$ blocks $IC_{5a}$ bias current from the pot wiper. Capacitor $C_{44}$ gives final d.c.-blocking to protect the following power amplifier.

Many components in this design are the same value; for example, wherever a sizable non-electrolytic is required, 470 nF could usually be made to work. This philosophy has to be abandoned in areas where critical parameters are set, such as the RIAA network and tone control stage.

## Supplying power

This is a conventional power supply using IC regulators. I strongly recommend that you use a toroidal mains transformer to minimise the a.c. magnetic field.

Supply rails have been increased from ±15 to ±18 V to maximise headroom. Nonetheless, 15 V regulators are specified as they are easy to obtain. Their output increased to 18 V by means of $R_{201,203}$, $Tr_{201}$ and $R_{202,204}$, $Tr_{202}$.

It is common to use a potential divider to 'stand-off' the regulator by a fixed proportion of the output voltage. In the improved version here, positive divider $R_{201,203}$ is buffered by emitter-follower $Q_{201}$. Thus $R_{201,203}$ can be higher in value – saving power – while $Tr_{201}$ absorbs the ill-defined quiescent current from the regulator COM pin.

## Choosing the right op-amps

Exotic and expensive op-amps will probably give a disappointing noise performance. The bipolar input of the *5534/32* is well matched to the medium-low impedances used in this preamplifier. For example, an *OP-27* might be expected to be quieter in the moving-magnet cartridge stage; but when measured, or calculated, it is 2 dB noisier.

## The performance

Figure 12 shows the THD of the flat moving-coil cartridge stage alone, at maximum gain. The rise at extreme if is due to the integrator time-constant. Figures 13 and 14 give the THD of the moving-magnet cartridge disc input and the entire rear section respectively. Levels involved are ten times those found in real use. Distortion is not a problem here.

Crosstalk performance attained depends very much on physical layout. Capacitive crosstalk can be minimised by spacing components well apart, or by simple screening. Resistive crosstalk depends on the thickness of the various ground paths.

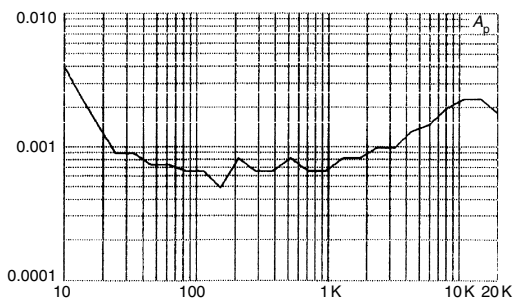



**Figure 12** *THD of the moving coil stage alone, at* 2.2 $V_{rms}$ *output. Measurement bandwidth* 30 kHz.

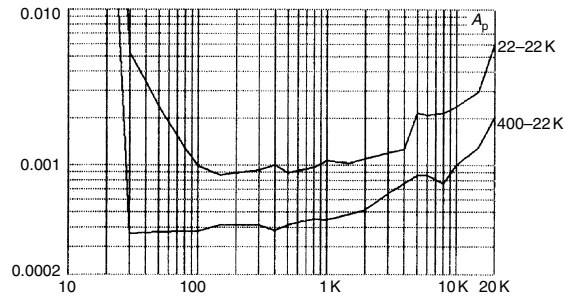



**Figure 13** *THD of disc input stage in moving-magnet mode, at* $8V_{rms}$ *output. Bandwidth* 22–22 kHz *upper trace and* 400–22 kHz *lower trace, which gives a more valid result as magnetic hum is excluded. Distortion is very low, but rises at h.f. due to increasing loading.*

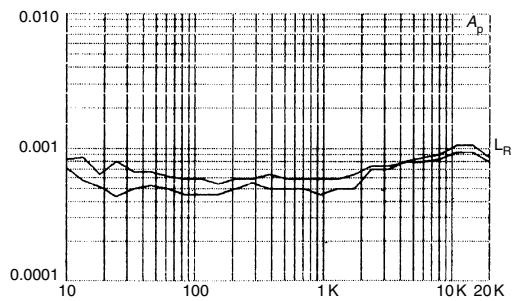



**Figure 14** *THD of rest of preamplifier at* $8V_{rms}$ *in and out, i.e. volume control set for unity gain. Tone control set flat, bandwidth* 80 kHz. *Distortion is below 0.001%.*

It would be desirable to specify a grounding topology for optimal results, but this is not so easy. I found that the more tightly the various grounds are tied together with heavy conductors, the better the crosstalk performance. There seemed little scope for subtlety.

As with noise performance, the results depend somewhat on control settings, but under most conditions the prototype gave about −100 dB flat across 20 Hz–20 kHz, with noise contributing to the reading. This was not hard to achieve.

## The preamplifier in perspective

In determining what (if anything) has been achieved by this design, we must see if it is capable of any further improvement.

- The moving-magnet stage input noise performance is limited by the electrical characteristics of the cartridge and its loading needs.
- Making the RIAA any more accurate will be expensive.
- Increasing disc input headroom would require the use of higher supply rails, demanding discrete amplifier stages.

Having gone to some effort to make the preamplifier as noise-free and transparent as possible, we should ask how it compares with other parts of the system. The standard Blameless Class B power amplifier[4] output noise is −93.5 dBu, and the Trimodal[5] with the low-impedance feedback network reduces this to −95.4 dBu. In both cases the source impedance is 50 Ω.

Both amplifiers have a closed-loop gain of +27.2 dB, and so the equivalent input noise (EIN) is −120.7 and −122.6 dBu respectively. This can be compared with the source-resistance Johnson noise of 50 Ω, which is −135.2 dBu. The best power-amp noise figure is therefore 12.6 dB, which is some way short of perfection.

In contrast, the noise output from the preamplifier is never less than −114.5 dBu with the volume control at zero. Even in this rather useless condition, the preamplifier increases the total noise output, as it produces 8 dB more than the Trimodal power amplifier input noise. At mid-volume (in-line mode) the preamplifier noise is −105.3 dBu, which is 17 dB worse than the power-amp; clearly as far as preamp design is concerned, history has not yet ended.

Even so, serious thought has been given to whether this may be the quietest preamp yet built. Comments and opinions on this are invited.

## References

1. Self, D. 'A Precision Preamplifier', *Wireless World*, October 1983, p 31.
2. Self, D. 'Ultra-Low-Noise Amplifiers and Granularity Distortion', *Journ AES*, November 1987, pp 907–915.
3. Baxandall, P. 'Audio Gain Controls', *Wireless World*, November 1980, pp 79–81.
4. Self, D. 'Distortion in Power Amplifiers: Part 2', *Electronics World*, September 1993, p 736.
5. Self, D. 'Trimodal Audio Power', *Electronics World*, June 1995, p 462.

# 7 Overload matters

*February 1997*

There are no gain controls on RIAA inputs, so the overload margin, or headroom, is of considerable importance. The issue can get a bit involved, as the frequency-dependant gain is further complicated by a heavy frequency-dependant load in the shape of the feedback network. This heavy loading was a major cause of distortion and headroom-limitation in conventional RIAA stages that had emitter-follower outputs with highly asymmetrical drive capabilities, and for some reason it took the industry a very long time to wake up to this. Once more the 5532 op-amp solved *that* problem.

There are also interesting limitations to the levels which stylus-in-vinyl technology can generate. This chapter tries to sort out the issues involved in maximising disc input headroom.

There was no room in my Preamp '96 article for a proper discussion of the overload behaviour of RIAA preamp stages.[1] Like noise performance, the issue is considerably complicated by both cartridge characteristics and the RIAA equalisation.

There are some inflexible limits to the signal level possible on vinyl disc, and they impose maxima on the signal that a cartridge can reproduce. The absolute value of these limits may not be precisely defined, but they set the way in which maximum levels vary with frequency, and this is perhaps of even greater importance.

Figure 1(a) shows the physical groove amplitudes that can be put onto a disc. From subsonic up to about 1 kHz, groove amplitude is the constraint. If the sideways excursion is too great, the spacing will need to be increased to prevent one groove breaking into another, and playing time will be reduced. From about 1 kHz to ultrasonic, the limit is groove velocity rather than amplitude. If the cutter head tries to move sideways too quickly compared with its forward motion, the back facets of the cutter destroy the groove that has just been cut by the forward edges.

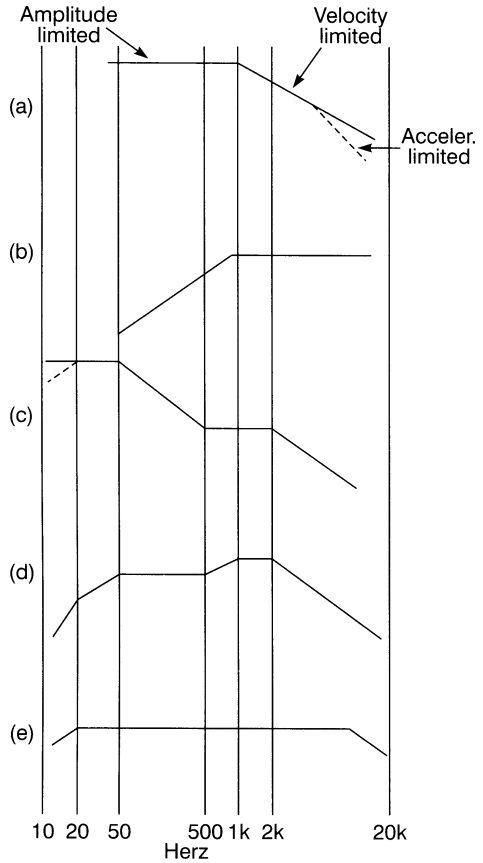



**Figure 1** *(a) Restrictions on the level put onto a vinyl disc. The extra limit of groove curvature – stylus acceleration – is shown dotted. (b) Response of a moving-coil or moving-magnet cartridge to a signal following the maximum contour in Figure 1(a). (c) The RIAA replay curve. The IEC amendment is an extra roll-off at low frequency, shown dotted. (d) The combination of (b) and (c). (e) RIAA preamp output limitations. The high-frequency restriction is very common and is often much worse in discrete preamplifier stages with poor load-driving capabilities.*

At replay time, there is a third restriction – that of stylus acceleration or, to put it another way, groove curvature. This sets a limit on how well a stylus of a given size can track the groove. Allowing for this at cutting time puts an extra limit on signal level, shown by the dotted line in Figure 1(a).

The severity of this restriction depends on the stylus shape. An old-fashioned spherical type with a tip diameter of 0.0007 in requires a roll-off of maximum levels from 2 kHz, while a relatively modern elliptical type with 0.0002 in effective diameter postpones the problem to about 8 kHz.[2]

Thus there are at least three limits on the signal level. The distribution of amplitude with frequency for the original signal is unlikely to mimic this, because there is almost always more energy at 1.f. than h.f. Therefore the h.f. can be boosted to overcome surface noise without overload problems, and this is done by applying the inverse of the familiar RIAA replay equalisation.

Moving-magnet and moving-coil cartridges both operate by the relative motion of conductors and magnetic field, so the voltage produced is proportional to rate of change of flux. The cartridge is sensitive to velocity rather than to amplitude (and so sensitivity is always expressed in millivolts per cm/s) and this gives a frequency response rising steadily at 6 dB/octave across the whole audio band. Therefore, a maximal signal from disc (Figure 1(a) would give a cartridge output like Figure 1(b) – i.e., 1(a) tilted upwards.

Figure 1(c) shows the RIAA replay equalisation curve. The shelf in the middle corresponds with 1(a), while an extra time constant at 50 Hz limits the amount of if boost applied to warps and rumbles. The 'IEC amendment' is an extra roll-off at 20 Hz, (shown dotted) to further reduce subsonics. When RIAA equalisation 1(c) is applied to cartridge output 1(b), the result will look like Figure 1(d), with the maximum amplitudes occurring around 1–2 kHz.

Clearly, the overload performance of an RIAA input can only be assessed by driving it with an inverse-RIAA equalised signal, rising at 6 dB/octave except around the middle shelf. My Precision preamp '96 has an input overload margin referred to 5 mV r.m.s. of 36 dB across most of the audio band, i.e., 315 mV r.m.s. at 1 kHz. The margin is still 36 dB at 100 Hz, but due to the RIAA low-frequency boost this is only 30 mV r.m.s. in absolute terms.

The final complications is that preamplifier output capability almost always varies with frequency. In Preamp '96, the effects have been kept small. The output overload margin voltage – and hence input margin – falls to +33 dB at 20 kHz. This is due to the heavy capacitive loading of both the main RIAA feedback path and the pole-correcting *RC* network ($R_{24,25}$ and $C_{20}$). This could be eliminated by using an op-amp with greater load-driving capabilities, if you can find one with the low noise of a *5534*.

The overload capability of Preamp 96 is also reduced to 31 dB in the bottom octave 10–20 Hz, because the IEC amendment is implemented in the second stage. The 1.f. signal is fully amplified by the first stage, then attenuated by the deliberately slow initial roll-off of the subsonic filter.

Such audio impropriety always carries a penalty in headroom as the signal will clip before it is attenuated. This is the price paid for an accurate IEC amendment set by polyester caps in the second stage, as opposed to the usual method of putting a small electrolytic in the first-stage feedback path, rather than the 220 μF used. Alternative input architectures that put

flat amplification before an RIAA stage suffer much more severely from this kind of headroom restriction.[3]

These extra preamp limitations on output level are shown at Figure 1(e), and, comparing 1(d), it appears they are almost irrelevant because of the falloff in possible input levels at each end of the audio band.

## References

1. Self, D. 'Precision preamplifier '96', *Electronic World*, July/August and September 1996.
2. Holman, T. 'Dynamic range requirements of phonographic preamplifiers', *Audio*, July 1977, p 74.
3. Self, D. 'Precision Preamplifier '96', *Electronics World*, July/August 1996, p 543.

# 8 A balanced view, Part I

*April 1997*

People often get confused when grappling with the many kinds of balanced, unbalanced, and ground-cancelling inputs and outputs; given the many permutations and variations this is not surprising. Balanced or quasi-balanced connections have been used for many years in professional circles, and are now slowly but apparently steadily increasing in popularity in the more expensive reaches of the hi-fi market. In this area the outputs are usually balanced by driving the cold line with an inverter rather than going for the full quasi-balanced version, and very sensibly so. The input amplifiers are usually the classic one-op-amp differential stage, though now and then some very complicated instrumentation-amp circuits have been used. For some reason no one seems to have picked up on the great merits of ground-cancelling outputs, possibly because of the difficulty of explaining the concept to the paying customers.

These two articles are an attempt to explain the operation of various kinds of inputs and outputs, some obvious, some almost unknown, with guidance on joining them together effectively.

Balanced inputs and outputs have been used for many years in professional audio, but profound misconceptions about their operation and effectiveness still survive. Balanced operation is also making a slow but steady advance into top-end hi-fi, where its unfamiliarity can lead to further misunderstandings. As with most topics in audio technology, the conventional wisdom is often wrong.

A practical balanced interconnect is not always wholly straightforward. Some new variations on input and output stages have emerged relatively recently. For example, a 'ground-cancelling' output is not balanced at all, but actually has one output terminal configured as an input. This can come as a surprise to the unwary.

Despite its non-balanced nature, such a ground-cancelling output can render a ground loop innocuous even when driving an unbalanced input. But even an audio professional could be forgiven for being unsure if it still works when it is driving a balanced input. The answer – which in fact is yes – is explained in a second article on this subject, details of which are given later.

## Electronic versus transformer balancing

Electronic balancing has many advantages. These include low cost, low size and weight, superior frequency and transient response, and no problems with low-frequency linearity. While it is sometimes regarded as a second-best, it is more than adequate for hi-fi and most professional applications.

### To balance . . .

Balancing offers the following advantages

- Discriminates against noise and crosstalk.
- A balanced interconnect – with a true balanced output – allows 6 dB more signal level on the line.
- Breaks ground-loops, so that people are not tempted to start 'lifting grounds'. This is only acceptable if the equipment has a dedicated ground-lift switch, that leaves the metalwork firmly connected to mains safety earth. In the absence of this facility, the optimistic will remove the mains earth – which is not quite so easy now that moulded plugs are standard – and this practice must be roundly condemned as dangerous.

### . . . or not to balance

Balancing also brings with it the following disadvantages.

- Balanced connections are unlikely to provide much protection against r.f. ingress. Both sides of the balanced input would have to demodulate the r.f. with exactly the same effectiveness for common-mode cancellation to occur. This is not very likely.
- There are more possibilities for error when wiring up. For example, it is easy to introduce an unwanted phase inversion by confusing hot and cold in a connector. This can go undiscovered for some time. The same mistake on an unbalanced system interrupts the audio completely.

Transformer balancing has some advantages of its own – particularly for work in very hostile r.f./e.m.c. environments – but many serious drawbacks. The advantages are that transformers are electrically bullet-proof, retain their common-mode rejection ratio performance forever, and consume no power even at high signal levels. Unfortunately they also generate low frequency distortion, and have high frequency response problems due to leakage reactance and distributed capacitance. Transformers are also heavy and expensive.

The first two objections can be surmounted – given enough extra electronic circuitry – but the last two cannot. Transformer balancing is therefore rare, even in professional audio, and is only dealt with briefly here.

## Balancing basics

Balanced connections in an audio system are designed to reject both external noise, from power wiring etc., and also internal crosstalk from adjacent signal cables.

The basic principle of balanced interconnection is to get the signal you want by subtraction, using a three-wire connection. In many cases, one signal wire – the hot or in phase conductor – senses the actual output of the sending unit. The other, the cold or phase-inverted, senses the unit's output-socket ground, and the difference between them gives the wanted signal.

Any noise voltages that appear identically on both lines, i.e. common-mode signals, are in theory completely cancelled by the subtraction. In real life, the subtraction falls short of perfection, as the gains via the hot and cold inputs will not be precisely the same. The degree of discrimination actually achieved is called the common-mode rejection ratio (cmrr).

The terms hot and cold for in-phase and out-of-phase respectively, are used throughout this article for brevity.

While two wires carry the signal, the third is the ground wire which has the dual duty of both joining the grounds of the interconnected equipment, and electrostatically screening the two signal wires by being in some way wrapped around them. The 'wrapping around' can mean:

- A lapped screen, with wires laid parallel to the central signal conductor. The screening coverage is not perfect, and can be badly degraded as it tends to open up on the outside of cable bends.
- A braided screen around the central signal wires. This is more expensive, but opens up less when the wire is bent. Screening is not 100%, but certainly better than lapped screen.
- An overlapping foil screen, with the ground wire – called the drain wire in this context for some reason – running down the inside of the foil

and in electrical contact with it. This is usually the most effective as the foil cannot open up on the outside of bends, and should give perfect electrostatic screening. However, the higher resistance of aluminium foil compared with copper braid means that r.f. screening may be worse.

## Electrical noise

Noise gets into signal cables in three major ways:

### *Electrostatic coupling*

An interfering signal with significant voltage amplitude couples directly to the inner signal line, through stray capacitance. The situation is shown in Figure 1, with *C*, *C* representing the stray capacitance between imperfectly-screened conductors; this will be a fraction of a picofarad in most circumstances. This coupling is unlikely to be a problem in hi-fi systems, but can be serious in studio installations with unrelated signals going down the same ducting.

The two main lines of defence against electrostatic coupling are effective screening and low-impedance drive. An overlapping foil screen – such as used on Belden microphone cable – provides complete protection. Driving the line from a low impedance, of the order of 100 Ω or less, means that the interfering signal, having passed through a very small capacitance, is

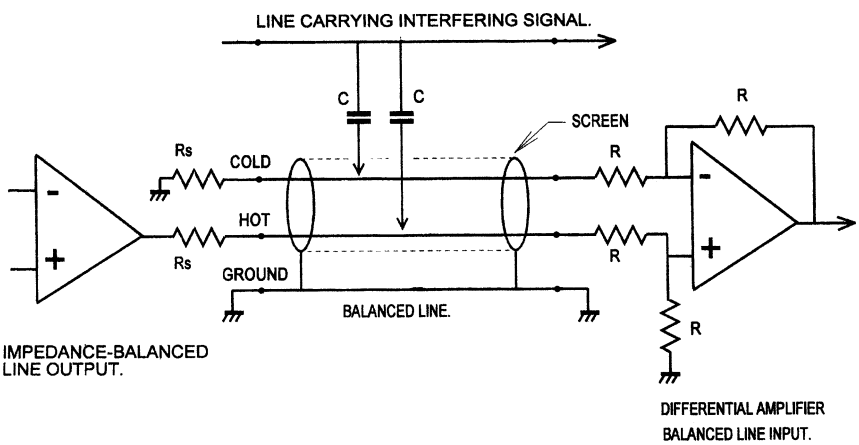



**Figure 1** *Electrostatic coupling into a signal cable,* $R_s$ *is* 100 Ω *and* R *is* 10 kΩ. *The second* $R_s$ *to ground in the cold output line makes it an impedance balanced output.*

a very small current and cannot develop much voltage across such a low impedance.

For the best results, the impedance must remain low up to as high as frequency as possible; this can be problem as op-amps invariably have a feedback factor that begins to fall from a low, and possibly sub-audio frequency, and this makes the output impedance rise with frequency.

From the point of view of electrostatic screening alone, the screen does not need to be grounded at both ends, or form part of a circuit.[1] It must of course be grounded at some point.

Electrostatic coupling falls off with the square of distance. Rearranging the cable-run away from the source of interference is more practical and more effective than trying to rely on very good common-mode rejection.

## *Magnetic coupling*

An e.m.f., $V_m$, is induced in both signal conductors and the screen, Figure 2. According to some writers, the screen current must be allowed to flow freely, or its magnetic field will not cancel out the field acting on the signal conductors. Therefore the screen should be grounded at both ends, to form a circuit.[2]

In practice, the field cancellation will be far from perfect. Most reliance is placed on the common-mode rejection of the balanced system, to cancel out the hopefully equal voltages $V_m$ induced in the two signal wires. The need to ground both ends for magnetic rejection is not a restriction, as it

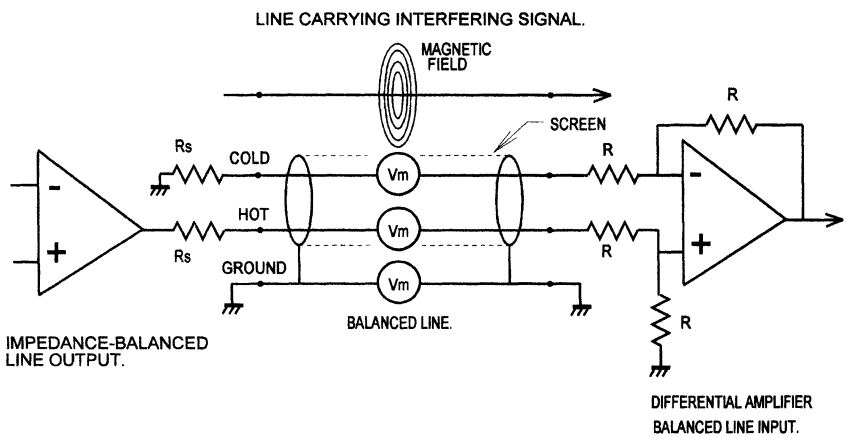



**Figure 2** *Magnetic coupling into a signal cable, represented by notional voltage-sources $V_m$.*

will emerge that there are other good reasons why the screens should be grounded at both ends of a cable.

In critical situations, the equality of these voltages is maximised by minimising the loop area between the two signal wires, usually by twisting them tightly together. In practice most audio cables have parallel rather than twisted signal conductors, and this seems adequate most of the time.

Magnetic coupling falls off with the square of distance, so rearranging the cable-run away from the source of magnetic field is usually all that is required. It is unusual for it to present serious difficulties in a domestic environment.

### *Common-impedance coupling*

Ground voltages coupled in through the common ground impedance; often called 'common-impedance coupling' in the literature.[3] This is the root of most ground loop problems. In Figure 3 the equipment safety grounds cause a loop ABCD; the mere existence of a loop in itself does no harm, but it is invariably immersed in a 50 Hz magnetic field that will induce mains-frequency current plus odd harmonics into it. This current produces a voltage drop down the non-negligible ground-wire resistance, and this once again effectively appears as a voltage source in each of the two signal lines. Since the cmrr is finite a proportion of this voltage will appear to be differential signal, and will be reproduced as such.

A common source of ground-loop current is the connection of a system to two different 'grounds' that are not actually at the same a.c. potential. The classic example of this is the addition of a 'technical ground' such as a buried copper rod to a grounding system which is already connected to 'mains ground' at the power distribution board. In most countries this

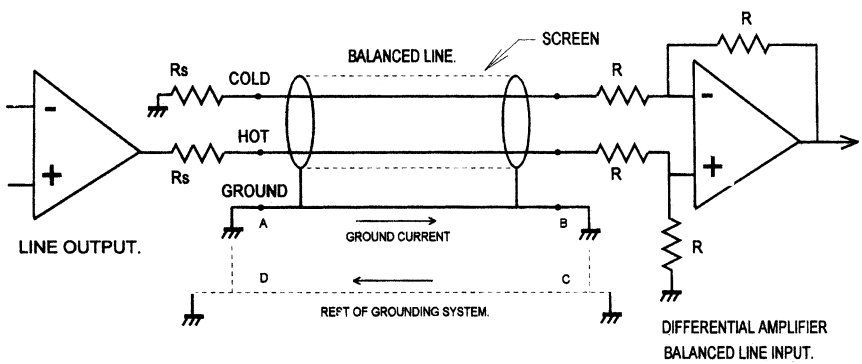



**Figure 3** *Ground-voltages coupling into a signal cable. The ground voltage between A and B is due to ground currents flowing around ABCD.*

'mains ground' is actually the neutral conductor, which is only grounded at the remote transformer substation. The voltage-drop down the neutral therefore appears between 'technical ground' and 'mains ground' causing large currents to flow through ground wires.

A similar situation can occur when water-pipes are connected to 'mains ground' except that interference is not usually by a common ground impedance; however the unwanted currents flowing in the pipework generate magnetic fields that may either create ground loops by induction, or interfere directly with equipment such as mixing consoles.

In practice, ground voltages cause a far greater number of noise problems than the other mechanisms, in both hi-fi and professional situations.

Even there is no common-impedance coupling, ground currents may still enter the signal circuit by transformer action. An example of such a situation is where the balanced line is fully floating and not galvanically connected to ground – which is only possible with a transformer-to-transformer connection.

The shield wire or foil acts as a transformer primary while the signal lines act as secondaries; if the magnetic field from the shield wire is not exactly uniform, then a differential noise voltage appears across the signal pair and is amplified as if it were a genuine signal. This effect is often called shield-current-induced-noise, or SCIN, and cables vary in their susceptibility to it according to the details of their construction.[4]

Fortunately the level of this effect is below the noise-floor in most circumstances and with most cables, for once a differential-mode signal has been induced in the signal lines, there is no way to discriminate against it.

From this summary I deduce there are two principle effects to guard against; electrostatic coupling, and the intrusion of unwanted voltages from either magnetic coupling or ground-loop currents.

Electrostatic interference can be represented by notional current sources connected to both signal lines; these will only be effectively cancelled if the line impedances to ground are the same, as well as the basic cmrr being high. The likely levels of electrostatic interference current in practice are difficult to guess, so the figures I give in the second article are calculated from applying 1 mA to each line; this would be very severe crosstalk, but it does allow convenient relative judgements to be made.

Magnetic and ground-voltage interference can be represented by notional voltage-sources inserted in both signal lines and the ground wire; these are not line-impedance sensitive and their rejection depends only on the basic cmrr, as measured with low-impedance drive to each input. Similarly ground-voltage interference can be represented by a voltage-source in the ground wire only.

Both input and output are voltages so the cmrr can be quoted simply as a ratio in decibels, without specifying any level.

## Line outputs

A line output is expected to be able to drive significant loads, partly because of a purely historical requirement to drive 600 Ω, and partly to allow the parallel feed of several destinations. Another requirement is a low source impedance −100 Ω or less – to make the signal robust against capacitive crosstalk, etc.

There are many line output and input arrangements possible, and the results of the various permutations of connection are not always entirely obvious. An examination of the output types in use yields List 1.

### List 1. Line output arrangements

- Unbalanced output
- Impedance-balanced output
- Ground-cancelling output, or ground-compensated output
- Balanced output
- Quasi-floating output
- True floating transformer output

### *Unbalanced output*

There are only two physical output terminals – signal and ground, Figure 4(a). A third terminal is implied in Figure 4(a), emphasising that it is always possible to connect the cold wire in the cable to the ground at the transmitting (output) end.

The output amplifier is almost always buffered from the line shunt-capacitance by a resistor $R_s$ in the range 33 to 100 Ω, to ensure stability. This unbalances the line impedances. If the output resistance is taken as 100 Ω worst-case, and the cold line is simply grounded as in Figure 4(a),

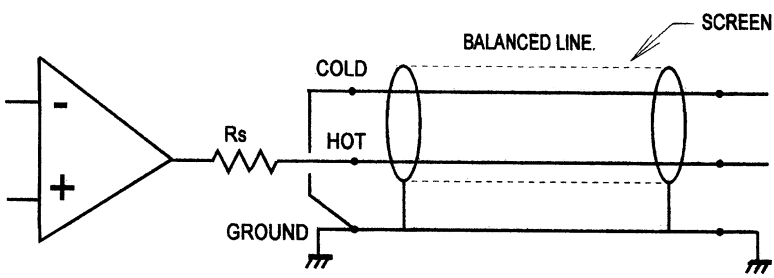



**Figure 4(a)** *An unbalanced line output. The cold output – if it exists at all – is connected directly to ground.*

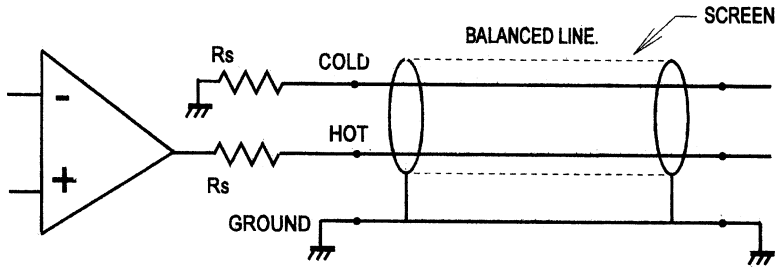



**Figure 4(b)** *An impedance balanced output. The cold output is connected to ground through a second* $R_s$ *of identical value.*

then the presence of $R_s$ degrades the common-mode rejection ratio to −46 dB, even if the balanced input at the other end of the cable has perfectly matched resistors.

### *Impedance balanced output*

There are now three physical terminals, hot, cold, and ground, Figure 4(b). The cold terminal is neither an input nor an output, but a resistive termination with the same resistance $R_s$ as the hot terminal output impedance. This type of output is intended for use with receiving equipment having balanced inputs. The presence of the second $R_s$ terminated to output ground makes the impedance on each signal line almost exactly the same – apart from op-amp output impedance limitations – so that good rejection is achieved for both common-mode ground voltages and electrostatic interference.

If an unbalanced input is being driven, the cold terminal on the transmitting (output) equipment can be either shorted to ground locally or left open-circuit without serious consequences. Either way all the benefits of balancing are lost.

The use of the word 'balanced' is unfortunate as this implies anti-phase outputs, which are not present.

### *Ground-cancelling output*

Also called a ground-compensated output, this arrangement is shown in Figure 5(a).

This allows ground voltages to be cancelled out even if the receiving equipment has an unbalanced input. It prevents any possibility of creating a phase error by miswiring. It separates the wanted signal from the unwanted by addition at the output end of the link, rather than by subtraction at the input end.

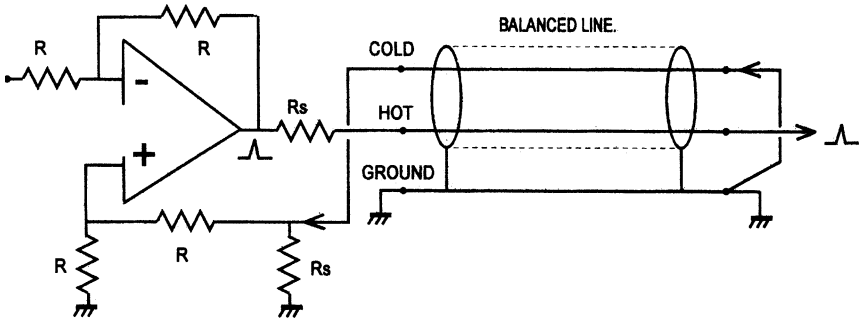



**Figure 5(a)** *A ground-cancelling output, with a unity-gain path from the cold terminal to the hot output. Once more a second* $R_s$ *balances the line impedances.*

If the receiving equipment ground differs in voltage from the sending ground, then this difference is added to the output so that the signal reaching the receiving equipment has the same voltage superimposed upon it. Input and ground therefore move together and there is no net input signal, subject to the usual resistor tolerances.

The cold pin of the output socket is now an input, and must have a unitygain path summing into the main signal-output going to the hot output pin. It usually has a very low input impedance equal to the hot terminal output impedance.

It is unfamiliar to most people to have the cold pin of an output socket as a low impedance input, and this can cause problems. Shorting it locally to ground merely converts the output to a standard unbalanced type. If the cold input is left unconnected then there should be only a very small noise degradation due to the very low input impedance of $R_s$.

Ground-cancelling outputs would appear to be very suitable for hi-fi use, as they are an economical way of making ground-loops innocuous. However, I am not aware that they have ever been used in this field.

## *Balanced output*

The cold terminal is now an active output, producing the same signal as the hot terminal but phase-inverted, Figure 5(b). This can be simply done by using an op-amp stage with a gain of minus one to invert the normal in-phase output. Phase spikes are shown on the diagram to emphasise these phase relationships.

The in-phase signal itself is not degraded by passing through an extra stage and this can be important in quality-critical designs. The inverting output must not be grounded; if not required it can simply be ignored.

Unlike quasi-floating outputs, it is not necessary to ground the cold pin to get the correct gain for unbalanced operation, and it must not be

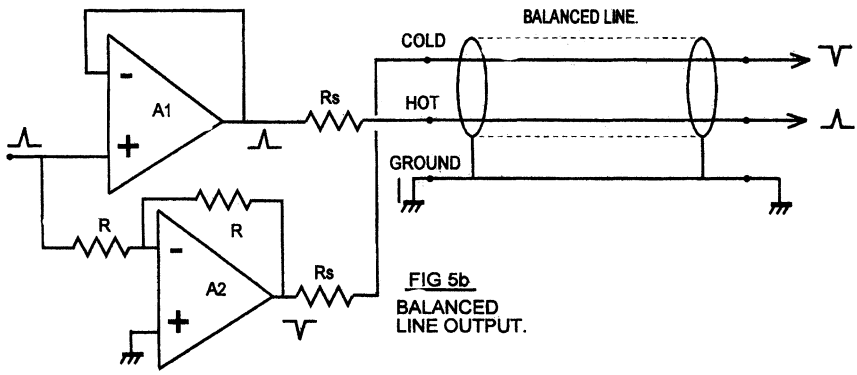



**Figure 5(b)** *A balanced output. A2 is a unity-gain inverter driving the cold output. Line impedances are balanced.*

grounded by mistake, because the inverting op-amp will then spend most of its time in current-limiting, probably injecting unpleasant distortion into the preamp grounding system, and possibly suffering unreliability. Both hot and cold outputs must have the same output impedance $R_s$ to keep the line impedances balanced.

A balanced output has the advantage that it is unlikely to crosstalk to other lines, even if they are unbalanced. This is because the current injected via the stray capacitance from each crosstalking line cancels at the receiving end.

Another advantage is that the total signal level on the line is increased by 6 dB, which can be valuable in difficult noise situations. All balanced outputs give the facility of correcting phase errors by deliberately swopping hot and cold outputs. This tactic is however a double-edged sword, because it is probably how the phase became wrong in the first place.

This form of balanced output is the norm in hi-fi balanced interconnection, but is less common in professional audio, where the quasifloating output gives more flexibility.

## *Quasi-floating output*

This kind of output, Figure 6, approximately simulates a floating transformer winding; if both hot and cold outputs are driving signal lines, then the outputs are balanced, as if a centre-tapped output transformer were being used.

If, however, the cold output is grounded, the hot output doubles in amplitude so the total level is unchanged. This condition is detected by the current-sensing feedback taken from the outside of the 75 Ω output

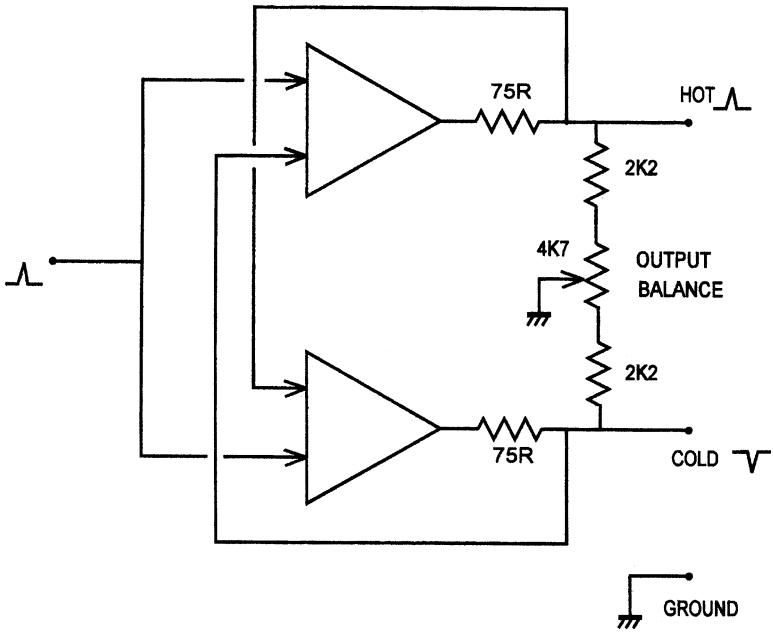



**Figure 6** *Simplified diagram of a quasi-floating balanced output, with its essential trim control for output symmetry.*

resistors. Current driven into the shorted cold output is automatically reduced to a low level that will not cause problems.

Similarly, if the hot output is grounded, the cold output doubles in amplitude and remains out of phase; the total hot-cold signal level is once more unchanged. This system has the advantage that it can give the same level into either a balanced or unbalanced input without rewiring connectors. 6 dB of headroom is however lost.

When an unbalanced input is being driven, the quasi-floating output can be wired to work as a ground-cancelling connection, with rejection of ground noise no less effective than the true balanced mode. This requires the cold output to be grounded at the remote (input) end of the cable. Under adverse conditions this might cause h.f. instability, but in general the approach is sound. If you are using exceptionally long cable, then it is wise to check that all is well.

If the cold output is grounded locally, i.e. at the sending end of the cable, then it works as a simple unbalanced output, with no noise rejection. When a quasi-floating output is used unbalanced, the cold leg must be grounded, or common-mode noise will degrade the noise floor by at least 10 dB, and there may be other problems. In both of the unbalanced cases the maximum signal possible on the line is reduced by 6 dB.

Quasi-floating outputs use a rather subtle circuit with an intimate mixture of positive and negative feedback of current and voltage. This performs the required function admirably; its only drawback is a tendency to accentuate circuit tolerances, and so a preset resistor is normally required to set the outputs for equal amplitude; the usual arrangement is shown in Figure 6.

If the balance preset is not correctly adjusted one side of the output will clip before the other and reduce the total output headroom. After factory setting this preset should not need to be touched unless the resistors in the circuit are replaced; changing the op-amp should make no difference.

The balancing network consists of a loading resistor to ground on each output; in this respect the output characteristics diverge from a true floating output, which would be completely isolated from ground. These loading resistors are lower than the input impedance of typical balanced inputs. So if simple differential amplifiers are used with unequal input impedances, (see the section on line inputs, below) the output balance is not significantly disturbed and clipping remains symmetrical on the hot and cold outputs.

Quasi-floating outputs are often simply referred to as 'balanced' or 'electronically-balanced', but this risks serious confusion as the true balanced output described earlier must be handled in a completely different way from quasi-floating.

### *True floating transformer output*

This can be implemented with a transformer if galvanic isolation from ground is required. The technique is rarely used.

The second article in this pair looks at line inputs in detail, examines what happens when the different kinds of input and output are connected together, and deals with the philosophy of audio system wiring.

## References

1. Williams, T. 'EMC for product designers', Newnes (Butterworth-Heinemann), p 1992, ISBN 0 7506 1264 9, p 176.
2. Williams, T. 'EMC for product designers', Newnes (Butterworth-Heinemann), p 1992, ISBN 0 7506 1264 9, p 173.
3. Muncy, N. 'Noise susceptibility in analog and digital signal processing systems', *JAES*, 3 (6), June 1995, p 447.
4. Muncy, N. 'Noise susceptibility in analog and digital signal processing systems', *JAES*, June 1995, p 441.

# 9 A balanced view, Part II

*May 1997*

There are only two kinds of input stage – unbalanced and balanced. For interconnection this is the primary distinction. Apart from balancing requirements, a line-level input, as opposed to a microphone input, is expected to have a reasonably high impedance to allow multiple connections to a single output.

Traditionally, a 'bridging impedance' – i.e. high enough to put negligible loading on historical 600 Ω lines – was 10 kΩ minimum. This is still appropriate for modern low-impedance outputs. However, a higher impedance of 100 kΩ or even more is desirable for interfacing to obsolete valve equipment, to avoid increased distortion and curtailed headroom.

Another common requirement is true variable gain at the balanced input, as putting the gain control further down the signal path means that it is impossible to prevent input amplifier overload. Thus you need a balanced stage that can attenuate as well as amplify, and this is where the circuit design starts to get interesting.

In the following circuitry, small capacitors often shunt the feedback elements to define bandwidth or ensure stability. These are omitted for clarity.

## *Unbalanced inputs*

These are straightforward; variable-gain series-feedback stages are easily configured as in Figure 1, providing a minimum gain of unity is acceptable; $R_2$ sets the gain law in the middle of the pot travel.

It is also simple to make a stage that attenuates as well as amplifies. But this implies a shunt-feedback configuration as in Figure 2, with a variable input impedance. The minimum input impedance $R_1$ cannot be much higher than 10 kΩ or resistor noise becomes excessive.

For a series-feedback stage, the input impedance can be made as high as desired by bootstrapping; an input resistance of 500 kΩ or greater is perfectly possible. This does *not* imply a poorer noise performance, as the noise depends on the source resistance and semi-conductor characteristics.

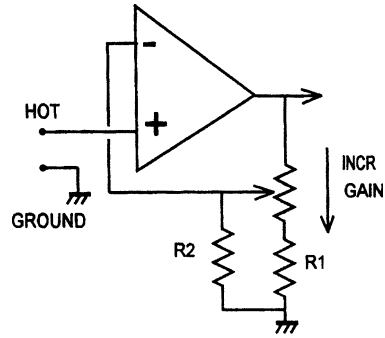



**Figure 1** *Variable-gain series-feedback unbalanced input stage. Resistor* $R_2$ *sets mid-position gain.*

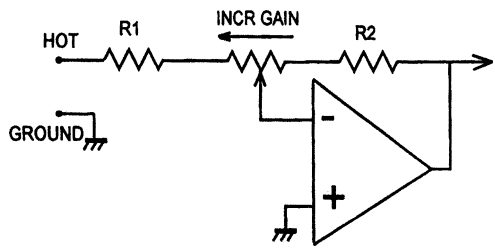



**Figure 2** *Shunt-feedback configuration, with a low and variable input impedance.*

To ram the point home, my own personal best is 1 GΩ, in a capacitor microphone head amplifier. Although the input impedance is many orders of magnitude greater than the 1 to 2 kΩ of a dynamic microphone preamp, the $E_{IN}$ is −110 dBu, i.e. only 18 dB worse.

Naturally, any unbalanced input can be made balanced or floating by adding a transformer.

## *Balanced inputs*

A standard one-op-amp differential input stage is shown in Figure 3. Unlike instrumentation work, a super-high cmrr is normally unnecessary. Ordinary 1% resistors and no trimming will not give cmrr better than 45 dB; however this is usually adequate for even high-quality audio work.

It is never acceptable to leave either input floating. This causes serious deterioration of noise, hum etc. Grounding the cold input locally to create an unbalanced input is quite alright, though naturally all the balanced noise rejection is lost.

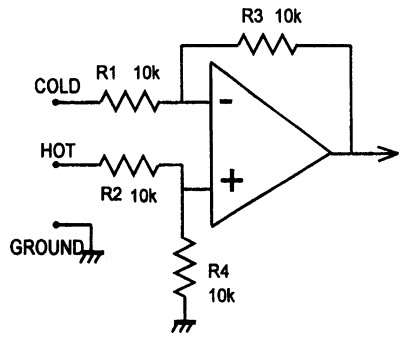



**Figure 3** *Standard one-op-amp differential amplifier, arranged for unity gain.*

The hot input can be locally grounded instead. In this case, the cold input is driven, to create a phase-inverting input that corrects a phase error elsewhere, but this is not good practice: the right thing to do is to sort out the original phase error.

## Balanced input technologies

There are many, many ways to make balanced or differential input amplifiers, and only the most important in audio are considered. These are:

- The standard differential amplifier
- Switched-gain balanced amp.
- Variable-gain balanced amp.
- The 'Superbal' amp.
- Hi-Z balanced amp.
- Microphone preamp plus attenuator
- Instrumentation amp.

### *Standard differential amplifier*

The standard one-op-amp differential amplifier is a very familiar circuit block, but its operation often appears somewhat mysterious. The version in Figure 3 has a gain of $R_3/R_1$ $(=R_4/R_2)$. It appears to present inherently unequal input impedances to the line; this has often been commented on[1] and some confusion has resulted.

The root of the problem is that a simple differential amplifier has interaction between the two inputs, so that the input impedance on the cold input depends strongly on the signal applied to the hot input. Since the only way to measure input impedance is to apply a signal and see how much

**Table 1** *Differential amplifier input impedances*

| *Case* | *Conditions* | *Hot l/p Z* | *Cold l/p Z* |
|---|---|---|---|
| 1 | Hot only driven | 20 kΩ | Grounded |
| 2 | Cold only driven | Grounded | 10 kΩ |
| 3 | Both driven balanced | 20 kΩ | 6.7 kΩ |
| 4 | Both driven cm, ie together | 20 kΩ | 20 kΩ |
| 5 | Both driven floating | 10 kΩ | 10 kΩ |

current flows into the input, it follows that the apparent input impedance on each leg varies according to the way the inputs are driven. If the amplifier is made with four 10 kΩ resistors, then the input impedances $Z$ are as in Table 1.

Some of these impedances are not exactly what you would expect. In Case 3, where the input is driven as from a transformer with its centre-tap grounded, the unequal input impedances are often claimed to 'unbalance the line'. However, since it is common-mode interference we are trying to reject, the common-mode impedance is what counts, and this is the same for both inputs.

The vital point is that the line output amplifier will have output impedances of 100 Ω or less, completely dominating the line impedance. These input impedance imbalances are therefore of little significance in practice; audio connections are not transmission lines (unless they are telephone circuits several miles long) so the input impedances do not have to provide a matched and balanced termination.

As the first thing the signal encounters is a 10 kΩ series resistor, the low impedance of 6.7 kΩ on the cold input sounds impossible. But the crucial point is that the hot input is driven simultaneously. As a result, the inverting op-amp input is moving in the opposite direction to the cold input, due to negative feedback, a sort of anti-bootstrapping that reduces the effective value of the 10 kΩ resistor to 6.7 kΩ.

The input impedances in this mode can be made equal by manipulating resistor values, but this makes the cm impedances (to ground) unequal, which seems more undesirable.

In Case 5, where the input is driven as from a floating transformer with any centre-tap unconnected, the impedances are nice and equal. They must be, because with a floating winding the same current must flow into each input. However, in this connection the line voltages are *not* equal and opposite: with a true floating transformer winding the hot input has all the signal voltage on it while the cold has none at all, due to the internal coupling of the balanced input amplifier.

This seemed very strange when it emerged from simulation, but a reality-check proved it true. The line has been completely unbalanced as

regards talking to other lines, although its own common-mode rejection remains good.

Even if perfectly matched resistors are assumed, the common-mode rejection ratio of this stage is not infinite; with a *TL072* it is about −90 dB, degrading from 100 Hz upwards, due to the limited open-loop gain of the opamp.

## *Switched-gain balanced amplifier*

The need for a balanced input stage with two switched gains crops up frequently. The classic application is a mixing desk to give optimum performance with both semi-professional (−7.8 dBu) and professional (+4 dBu) interface levels.

Since the nominal internal level of a mixer is usually in the range −4 to 0 dBu, the stage must be able to switch between amplifying and attenuating, maintaining good cmrr in both modes.

The obvious way to change gain is to switch both $R_{3,4}$ in Figure 3, but a neater technique is shown in Figure 4. Perhaps surprisingly, the gain of a differential amplifier can be manipulating by changing the drive to the feedback arm ($R_3$ etc.) only, without affecting the cmrr. The vital point is to keep the resistance of this arm the same, but drive it from a scaled version of the op-amp output.

Figure 4 uses the network $R_{5,6}$, which has the same 2 kΩ output impedance whether $R_4$ is switched to the output (low gain) or ground (high gain). For low gain, the feedback is not attenuated, but fed through $R_{5,6}$ in parallel.

For high gain, $R_{5,6}$ become a potential divider. Resistor $R_3$ is reduced by 2 kΩ to allow for the $R_{5,6}$ output impedance. The stage can attenuate as

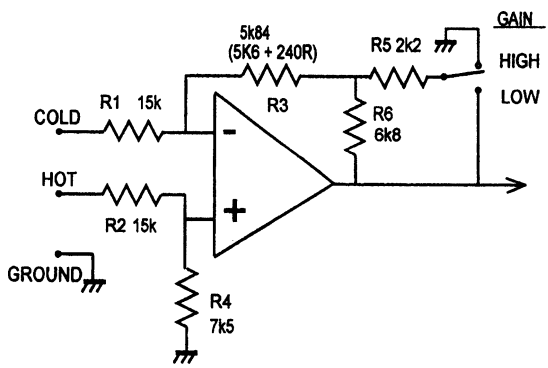



**Figure 4** *Switched-gain balanced input amplifier. The values shown give gains of* −6 dB *and* +6.2 dB*, for switching between pro and semi-pro interface levels.*

well as amplify if $R_1$ is greater than $R_3$, as shown here. The nominal output of the stage is assumed to be −2 dBu; the two gains are −6.0 and +6.2 dB.

The differential input impedance is 11.25 kΩ via the cold and 22.5 kΩ via the hot input. Common mode input impedance is 22.5 kΩ for both inputs.

### *Variable-gain balanced amplifier*

A variable-gain balanced input should have its gain control at the very first stage, so overload can always be avoided. Unfortunately, making a variable-gain differential stage is not so easy; dual potentiometers can be used to vary two of the resistances, but this is clumsy and will give shocking cmrr due to pot mismatching. For a stereo input the resulting four-gang potentiometer is unattractive.

The gain-control principle is essentially the same as for the switched-gain amplifier above. To the best of my knowledge, I invented both stages in the late seventies, but so often you eventually find out that you have re-invented instead; any comments welcome.

Feedback arm $R_3$ is of constant resistance, and is driven by voltage-follower $A_2$. This eliminates the variations in source impedance at the potentiometer wiper, which would badly degrade cmrr. As in Figure 1, $R_6$ modifies the gain law; however, the centre-detent gain may not be very accurate as it partly depends on the ratio of potentiometer track (often no better than ±10%, and sometimes worse) to 1% fixed resistors.

This stage is very useful as a general line input with an input sensitivity range of −20 to +10 dBu. For a nominal output of 0 dBu, the gain of Figure 5 is +20 to −10 dB, with $R_6$ chosen for 0 dB at the central wiper position.

An op-amp in a feedback path appears a dubious proposition for stability, but here, working as a voltage-follower, its bandwidth is maximised and in practice the circuit is dependably stable.

### *The 'Superbal' amplifier*

This configuration[2] gives much better input symmetry than the standard differential amplifier, Figure 6. The differential input impedance is exactly 10 kΩ via both hot and cold inputs. Common mode input impedance is 20 kΩ for both inputs. This configuration is less easy to modify for variable gain.

### *High-Z balanced amp*

High-impedance balanced inputs, above 10 kΩ, are useful for interfacing to valve equipment. Adding output cathode-followers to valve circuitry is

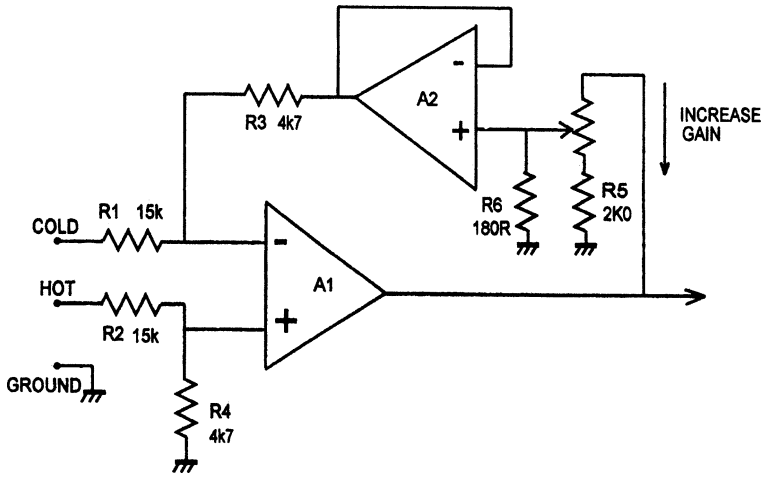



**Figure 5** *Variable-gain balanced input amplifier. Gain range is −10 to +20 dB. Resistor* $R_6$ *sets the mid-position gain.*

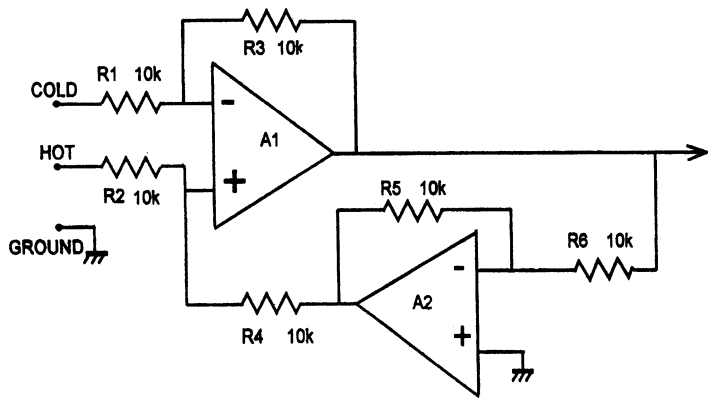



**Figure 6** *The 'Superbal' balanced input stage; input impedance on hot and cold are equal for both differential and common mode.*

expensive, and so the output is often taken directly from a gain-stage anode. Even a light loading of 10 kΩ may seriously compromise distortion and available output swing.

All of the balanced stages dealt with up to now have their input impedances determined by the values of input resistors etc., and these cannot be raised without degrading noise performance. Figure 7 shows

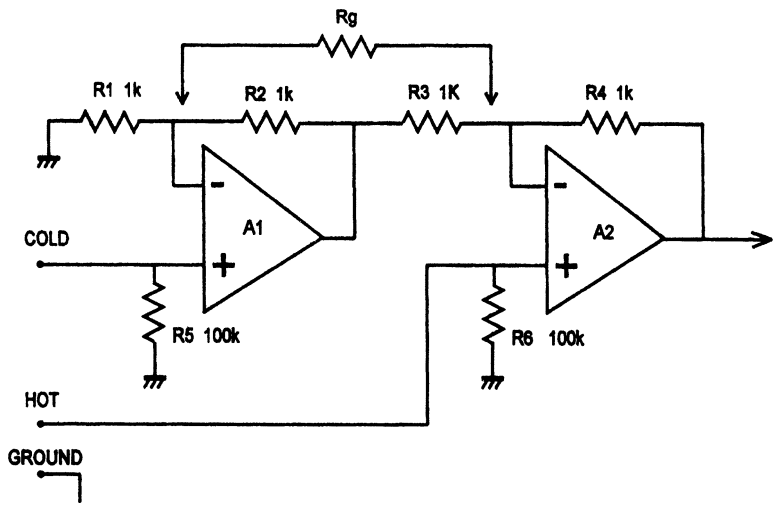



**Figure 7** *High-ïmpedance balanced input stage;* $R_5$ *and* $R_6$ *set input impedance, and can be much higher. Add* $R_8$ *to increase gain.*

one answer to this. The op-amp inputs have infinite impedance in audio terms, subject to the need for $R5$, $R6$ to bias the non-inverting inputs.[3]

Adding $R_g$ increases gain, but preserves balance. This configuration cannot be set to attenuate.

### *Microphone preamp with attenuator*

It is often convenient to-use a balanced microphone preamp as a line input by using a suitable balanced attenuator, typically 20 to 30 dB. The input impedance of the microphone input stage will be 1 to 2 kΩ for appropriate mic loading, and this constrains the resistor values possible.

Keeping the overall input impedance to at least 10 kΩ means that the divider impedance must be fairly high, with a lot of Johnson noise. As a result, the total noise performance is almost always inferior to a dedicated balanced line-input amplifier. Common-mode rejection ratio is determined by the attenuator tolerances and will probably be much inferior to the basic microphone amp, which usually relies on inherent differential action rather than component matching.

Figure 8(a) shows a bad way to do it; the differential signal is attenuated, but not the common-mode, so cmrr is degraded even if the resistors are accurate. Figure 8(b) attenuates differential and common-mode signals by the same amount, so cmrr is preserved, or at any rate no worse than resistor tolerances make it.

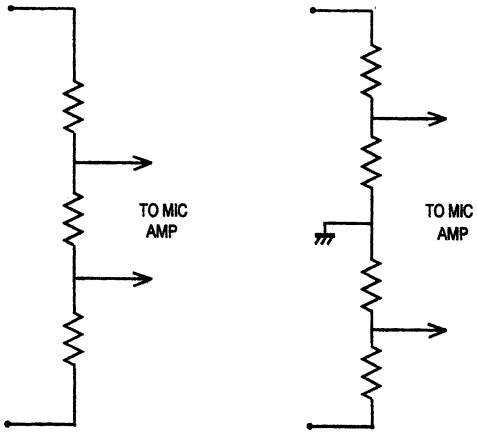



**Figure 8** *At (a), balanced attenuators convert a microphone preamp to line input. Circuit (b) is superior as both differential and commonmode signals are equally attenuated, so common-mode rejection is not degraded more than necessary.*

### *Instrumentation amplifier*

All the balanced inputs above depend on resistor matching to set the cmrr. In practice this means better than 45 dB is not obtainable without trimming. If a cmrr higher than this is essential, an IC instrumentation amplifier is a possibility.

Common-mode rejection ratio can be in the range 80 to 110 dB, without trimming or costly precision components. The IC tends to be expensive, due to low production volumes, and the gain is often limited in range and cannot usually be less than unity.

In audio work, cmrr of this order is rarely if ever required. If the interference is that serious, then it will be better to deal with the original source of the noise-rather than its effects.

## Input/output combinations

Taking five kinds of output – the rare case of floating output transformers being excluded – and the two kinds of input amplifier, there are ten possible combinations of connection. The discussion below assumes output $R_s$ is 100 Ω, and the differential input amplifier resistors $R$ are all 10 kΩ, as in Figure 3.

### *Unbalanced output to unbalanced input*

This is the basic connection. There is no rejection of ground noise (cmrr = unity) or electrostatic crosstalk; in the latter case the 1 mA notional

crosstalk signal yields a −20 dBv signal as the impedance to ground is very nearly 100 Ω.

### *Unbalanced output to balanced input*

Assuming the output ground is connected to the cold-line input, then in theory there is complete cancellation of ground voltages. This is true, *unless* the output has a series output resistor to buffer it from cable capacitance, – which is almost always the case – for this will unbalance the line.

If the output resistance is 100 Ω, and the cold line is simply grounded as in Figure 8(a), then $R_s$ degrades the cmrr to −46 dB even if the balanced input has exactly matched resistors.

The impedances on each line will be different, but not due to the asymmetrical input impedances of a simple differential amplifier; hot line impedance is dominated by the output resistance $R_s$ on the hot terminal (100 Ω) and the cold line impedance is zero as it is grounded at the output end. The rejection of capacitive crosstalk therefore depends on the unbalanced output impedance. It will be no better than for an unbalanced input, as for the unbalanced output to balanced input case. The main benefit of this connection is ground noise rejection, which solves the most common system problem.

### *Impedance-balance out to unbalanced in*

There is nothing to connect the output cold terminal to at the input end, and so this is the same as the ordinary unbalanced connection for the *unbalanced output to balanced input* configuration.

### *Impedance-balance out to balanced in*

In theory there is complete cancellation of both capacitive crosstalk and common-mode ground voltages, as the line impedances are now exactly equal.

Table 2 shows the improvement that impedance-balancing offers over a conventional unbalanced output, when driving a balanced input with exactly matched resistors.

The effect of tolerances in the impedance-balance resistor are also shown; the rejection of capactive crosstalk degrades as soon as the value moves away from the theoretical 100 Ω, but the cmrr actually has its point of perfect cancellation slightly displaced to about 98.5 Ω, due to second-order effects. This is of no consequence in practice.

**Table 2** *Impedance-balancing gives better CMRR than the conventional circuit*

| | *Capacitive* 1 mA | *CMRR* (dB) |
|---|---|---|
| Conventional | −20 dBv | −46 |
| Impedance-bal 99 Ω | −60 dBv | −101 |
| Impedance-bal 100 R | −∞ | −85 |
| Impedance-bal 101 R | −60 dBv | −79 |

### Ground-cancelling out to unbalanced in

There is complete cancellation of ground voltages, assuming the ground-cancel output has an accurate unity gain between its cold and hot terminals. This is a matter for the manufacturer.

Ground-cancelling in this way is a very efficient and cost-effective method of interconnection for all levels of equipment, but tends to be more common at the budget end of the market.

### Ground-cancelling out to balanced in

This combination needs a little thought. At first there appears to be a danger that the ground-noise voltage might be subtracted twice, which will of course be equivalent to putting it back in in anti-phase, gaining us nothing.

In fact this is not the case, though the cancellation accuracy is compromised compared with the impedance-balanced case; the common-mode rejection will not exceed 46 dB, even with perfect resistor matching throughout. Capacitive crosstalk is no better than for the 'Unbalanced output to balanced input' i.e. approximately −21 dB, which means virtually no rejection. However, this is rarely a problem in practice.

### Balanced output to unbalanced input

This is not a balanced interconnection. There is nowhere to connect the balanced cold output to; it must be left open-circuit, its signal unused, so there is a 6 dB loss of headroom in the link. The unbalanced input means the connection is unbalanced, and so there is no noise rejection.

### Balanced out to balanced in

A standard balanced system, that should give good rejection of ground noise and electrostatic crosstalk.

### *Quasi-floating out to unbalanced in*

Since the input is unbalanced, it is necessary to ground the cold side of the quasi-floating output. If this is done at the remote (input) end then the ground voltage drop is transferred to the hot output by the quasi-floating action, and the ground noise is cancelled in much the same way as a ground-cancelling output.

However, in some cases this ground connection must be local, i.e. at the output end of the cable, if doing it at the remote (input) end cause high-frequency instability in the quasi-floating output stage. This may happen with very long cables. Such local grounding rules out rejection of ground noise because there is no sensing of the ground voltage drop.

Perhaps the major disadvantage of quasi-floating outputs is the confusion they can cause. Even experienced engineers are liable to mistake them for balanced outputs, and so leave the cold terminal unconnected. This is not a good idea. Even if there are no problems with pickup of external interference on the unterminated cold output, this will cause a serious increase in internal noise. I believe it should be standard practice for such outputs to clearly marked as what they are.

### *Quasi-floating out to balanced in*

A standard balanced system, that should give good rejection of ground noise and electrostatic crosstalk.

The hot and cold output impedances are equal, and dominate the line impedance, so even if the line input impedances are unbalanced, there should also be good rejection of electrostatic crosstalk.

## Wiring philosophies

It has been assumed above that the ground wire is connected at both ends. This can cause various difficulties due to ground currents flowing through it.

For this reason some sound installations have relied on breaking the ground continuity at one end of each cable. This is called the one-end-only (OEO) rule.[4] It prevents ground currents flowing but usually leaves the system much more susceptible to r.f. demodulation. This is because the cable screen is floating at one end, and is now effectively a long antenna for ambient r.f.

There is also the difficulty that non-standard cables are required. A consistent rule as to which end of the cable has no ground connection must

be enforced. The OEO approach may be workable for a fixed installation that is rarely modified, but for touring sound reinforcement applications it is unworkable.

A compromise that has been found acceptable in some fixed installations is the use of 10 nF capacitors to ground the open screen end at r.f. only; however, the other problems remain.

The formal OEO approach must not be confused with 'lifting the ground' to cure a ground loop. Unbalanced equipment sometimes provides a ground-lift switch that separates audio signal ground from chassis safety ground; while this can sometimes be effective, it is not as satisfactory as balanced connections. Lifting the ground must *never* be done by removing the chassis safety earth; this removes all protection against a live conductor contacting the case and so creates a serious hazard. It is also in many cases illegal.

The best approach therefore appears to be grounding at both ends of the cable, and relying on the cmrr of the balanced connection to render ground currents innocuous. Ground currents of 100 mA appear to be fairly common; ground currents measured in amps have however been encountered in systems with serious errors.

A typical example is connecting incoming mains 'Earth' – which is actually 'Neutral' in many cases – to a technical ground such as a buried copper rod. Take a look the section headed 'Electrical Noise' in last month's article for more details.

Ground currents cause the worst problems when they flow not only through cable shields but also the internal signal wiring of equipment. For this reason the preferred practice is to terminate incoming ground wires to the chassis earth of the equipment. This keeps ground currents off pcbs, where the relatively high track resistances would cause bad common-impedance coupling, and preserves r.f. screening integrity.

Grounding is simplified for source equipment that has no other connections, such as double-insulated compact-disc players. These carry a 'square-in-a-square' symbol to denote higher standards of mains insulation, so that external metalwork need not be grounded for safety. Such equipment often has unbalanced outputs, and can usually be connected directly to an unbalanced input with good results, as there is no path for any ground currents to circulate in.

If a balanced input is used, then connecting the hot input to CD signal and the cold to CD 'ground' leaves the CD player ground floating, and this will seriously degrade hum and r.f. rejection. The real ground must be linked to CD player common.

I think this chapter shows that balanced line interconnections are rather more complex than is immediately obvious. Having said that, with a little caution they work very well indeed.

## References

1. Winder, S. 'What's the difference?' *Electronics World*, November 1995, p 989.
2. Ballou, G. (ed) 'Handbook For Sound Engineers – The Audio Cyclopaedia' Howard Sams 1987, ISBN 0–672–21983–2 (Superbal).
3. Smith, J.L. 'Modern operational circuit design', Wiley-Interscience, 1971, p 31 and 241.
4. Muncy, N. 'Noise susceptibility in analog and digital signal processing systems', *JAES*, Vol. 3 (6), June 1995, p 440.

# 10 High-quality compressor/limiter

## A variable law, low distortion attenuator incorporating second harmonic cancellation circuitry

*December 1975*

This was the first article I ever wrote for Wireless World, as it then was. By the time I was in my third year at Cambridge, I was deeply involved in audio electronics, and so when the time came to choose a subject for the design project that was an important part of the course, I went straight for audio. There were three main considerations: it was important to pick something that was virtually certain to have a successful outcome, I wanted to have fun doing it, and it should be suitable for publication in Wireless World. The project was done in the early months of 1973, and took a little to work its through to the top of the editorial pile; in those days the competition to publish in WW was fierce.

The technology described in this article is at first sight now somewhat obsolete, though in an area where directly heated triodes designed just after the First World war are prized, it is a bit difficult to come up with a working definition of 'obsolete'. At the time the junction FET was a great step forward in voltage-controllable gain; previous approaches included diode bridges, filament-lamp and photoresistor combinations, and ultrasonic chopping, none of which were very linear or very satisfactory. The FET VCA was reasonably linear if the signal levels were kept low, and beautifully simple in terms of circuitry, but the $V_{gs}$/channel resistance law was (and is) subject to wide

production spreads. With a feedback compressor/limiter such as the one described in the article, this was not a great drawback in a single-channel unit, as the sidechain generated whatever voltage was required for the attenuation needed. However, problems arose when two compressor channels were linked together for stereo operation, and a very tedious selection was required to get the suitably matched pairs of FETs.

Today most compressor/limiters are based on transconductance-based VCAs which have an absolutely predictable control-voltage law, and are very linear compared with FET VCAs. (A little distortion still remains, so the history of VCAs has not ended yet.) These ingenious devices had begun to appear in 1975, but they were very expensive indeed, and FET-based compressors remained popular for many years. The widest use of FETs was as gain-control elements in the Dolby cassette noise reduction system.

Compression and limiting play an increasingly important role in the resources of a modern sound studio. The conventional function of signal level control is to avoid overload, but it can be used in the realm of special effects. To date, however, relatively few designs for high-fidelity compressor/limiters have been published.

The main design problem is the voltage-controlled attenuator, v.c.a., which increases attenuation of the input signal in response to a voltage from a control loop as shown in Figure 1. In limiting, this circuit block continuously monitors the peak output level from the v.c.a. and acts to maintain an almost constant level if it exceeds a threshold value, or, in compression, allows it to increase more slowly than the v.c.a. input signal. This is illustrated in Figure 2, which shows the input-amplitude/output-amplitude characteristic for both compression and limiting. Note that limiting makes use of a much tighter slope to ensure that the output

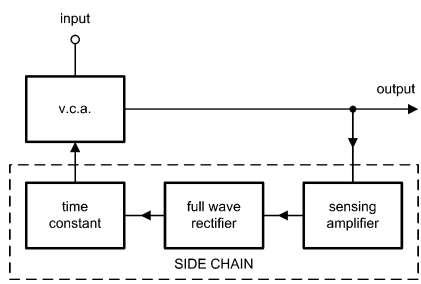



**Figure 1** *Voltage-controlled attenuator with d.c. control loop.*

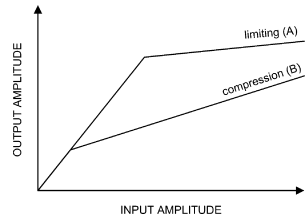



**Figure 2** *Amplitude characteristics for compression and limiting – the last mentioned uses an almost zero slope to prevent the output exceeding a preset level.*

voltage cannot exceed the chosen limit, and that the threshold (point of onset of attenuation) takes place at a higher level than for compression.

Traditionally, studio-quality compressor/limiters (as the two functions are so similar it is logical to produce a system that can be used for either compression or limiting) used one of two types of v.c.a. Either the audio signal was chopped at an ultrasonic frequency by a variable mark/space square wave – which requires complex circuitry and careful filtering of the audio output to avoid beats with tape-recorder bias frequencies – or it was attenuated by an electronic potential divider one arm of which was a photoresistor, the control signal being applied via a small filament bulb. The last-mentioned has disadvantages because photoresistors are non-linear devices, therefore noticeable distortion is introduced into the audio signal, and the thermal inertia of the bulb filament limits the speed of attenuation onset.

Most modern compression systems use field-effect transistor operated below pinch-off as a voltage-variable resistance in a potential divider. This technique has many advantages; it is a simple, cheap, and fast-acting configuration that can provide an attenuation variable between 0 and 45 dB. The only problem is that an FET is a square-law device, and tends to generate a level of second-harmonic distortion that increases rapidly with signal amplitude. A typical arrangement is shown in Figure 3 – $R_2$, $R_3$ and $C_2$ allow the source of the FET to be set at a d.c. level above ground, so that a control-voltage that moves positive with respect to ground can be used, to avoid level-shifting problems in the control loop. This d.c. level is isolated from the input and output by $C_1$.

The distortion introduced by this circuit is at its worst for the 6 dB attenuation condition, because at this point the drain-source resistance equals $R_1$, and the maximum power level exists in the FET. Table 1 shows the level of second-harmonic distortion introduced into a sine-wave signal of 100 mV r.m.s. amplitude, under the 6 dB attenuation condition for three different FET types. Measurements were made with a Marconi TF2330 wave analyser, higher-orders of harmonic distortion proved to be negligible amplitude in all cases. These measurements were made on one sample of

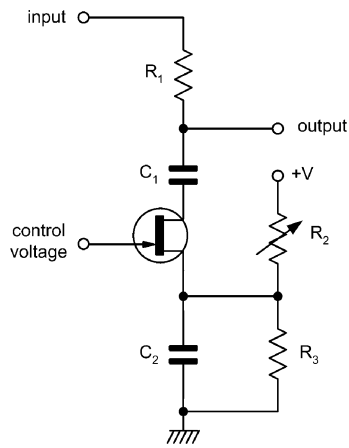



**Figure 3** *Basic v.c.a. circuit providing up to* 45 dB *of attenuation. This configuration introduces second-harmonic distortion which is greatest at* 6 dB *of attenuation.*

**Table 1** *Second-harmonic distortion level introduced into a sine-wave of 100 mV r.m.s.*

| *Device* | *2N3819* | *2N5457* | *2N5459* |
|---|---|---|---|
| 2 nd harmonic at −6 dB(%) | 13 | 10 | 8.9 |
| 2nd harmonic with cancellation (%) | 0.39 | 0.12 | 0.12 |
| attenuation shown (dB) | 2 | 10 | 2 |

each type of FET and, because production spreads are large, the results should be treated with some caution. However, it is clear that these levels of distortion are unacceptable for high-quality applications.

Fortunately, a technique exists for reducing FET distortion to manageable levels, if the control-voltage is applied to the FET gate and summed with a signal consisting of one-half the voltage from drain to source, then the distortion level is dramatically lowered. The configuration in Figure 4 shows a simple way of realising this; the signal fraction fed back is not critical and 10% resistors can be used for $R_4$ and $R_5$. Surprisingly, this distortion cancellation procedure leaves the attenuation/control-voltage characteristic almost unchanged. Table 1 shows the new maximum distortion values for 100 mV r.m.s. input. (Note that the maximum no longer occurs at 6 dB attenuation, but at a point that varies with the FET type, where cancellation is least effective.) From these results the 2N5457 and 2N5459 are superior, the 2N5459 was used in the final version of the v.c.a.

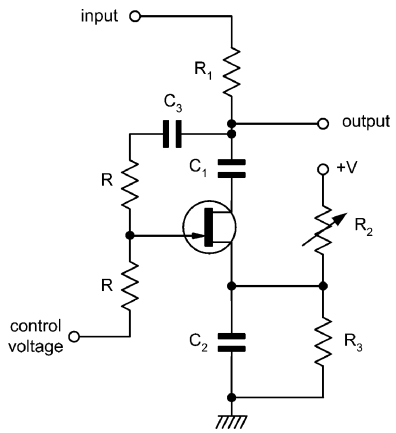



**Figure 4** *Standard circuit technique for reducing FET distortion by summing half of the drain/source voltage with the control voltage.*

To determine appropriate signal levels in the v.c.a., measurements were made of maximum distortion generated, i.e. the v.c.a. was set to 2 dB attenuation, against r.m.s. input voltage; results are shown in Table 2. The question now arises as to whether this distortion performance is adequate for a high-quality compressor/limiter. There is no general agreement as to the amount of second harmonic distortion that can be introduced into a program signal before it becomes aurally detectable, but 0.1% is a figure that is quoted. This means that the permissible input voltage to the v.c.a. would be restricted to below 100 mV r.m.s. In practice, however, the attenuation level will be constantly changing, and because distortion level peaks fairly sharply with attenuation change, this level of distortion will only be present for a very small percentage of the time. In any case, second harmonic distortion alone has a relatively low 'objectionability factor'. The proof of the pudding is in listening to the compressor output signal; inputs of music

**Table 2** *Maximum distortion generated by various input voltages at 2* dB *attenuation*

| *Input* (mV, r.m.s.) | *2nd harmonic (%)* |
|---|---|
| 20 | 0.005 |
| 50 | 0.10 |
| 100 | 0.12 |
| 200 | 0.19 |
| 500 | 0.34 |
| 1,000 | 0.56 |

around 200 mV r.m.s. produced no trace of audible distortion. (Good class A power amplifiers and headphones were used for monitoring).

The control loop consists of an amplifier which senses the v.c.a. output level. A full-wave rectification system is normal practice because program waveforms have positive and negative peaks that can vary by as much as 8 dB, and an 8 dB uncertainty in the output level is usually unacceptable. A time-constant arrangement is used with the rectification circuit to control the attack and decay rates.

The output sensing amplifier in the system is a non-inverting op-amp which allows a high input impedance because the output impedance of the v.c.a. stage reaches a maximum of about 39 kΩ at zero attenuation. The full-wave rectification system consists of a transistor phase-splitter driving two op-amp precision-rectifier stages in antiphase. The principle of a precision rectifier is illustrated in Figure 5. The rectifying element is placed in the feedback loop of an op-amp, so that the effect of the forward voltage drop on the output voltage is divided by the open-loop gain. During positive half-cycles, if the input voltage exceeds the d.c. level stored on the capacitor C, the op-amp output swings positive and C is charged through diode D until its stored voltage is equal to the input voltage. Thus C takes up a voltage across it equal to that of the positive peak of the input signal. During negative half-cycles, and while the input is less than the voltage on C during positive half-cycles, the op-amp saturates negatively and D remains firmly reverse-biased. Obviously this is only a half-wave rectification circuit, the full-wave version uses two of these driven in antiphase, and charging a common capacitor. A resistance through which the charging currents flow determines the attack time, and another in parallel with C defines the decay time-constant.

The complete circuit is shown in Figure 6. The v.c.a. is essentially as described above and the attenuation threshold is set by the variable resistance $R_2$. As the resistance is increased the level of control voltage required

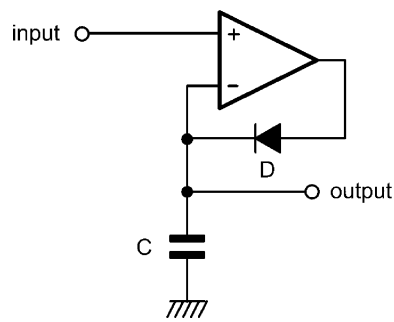



**Figure 5** *Basic precision rectifier circuit where the rectifying element is in the feedback loop of an op-amp.*

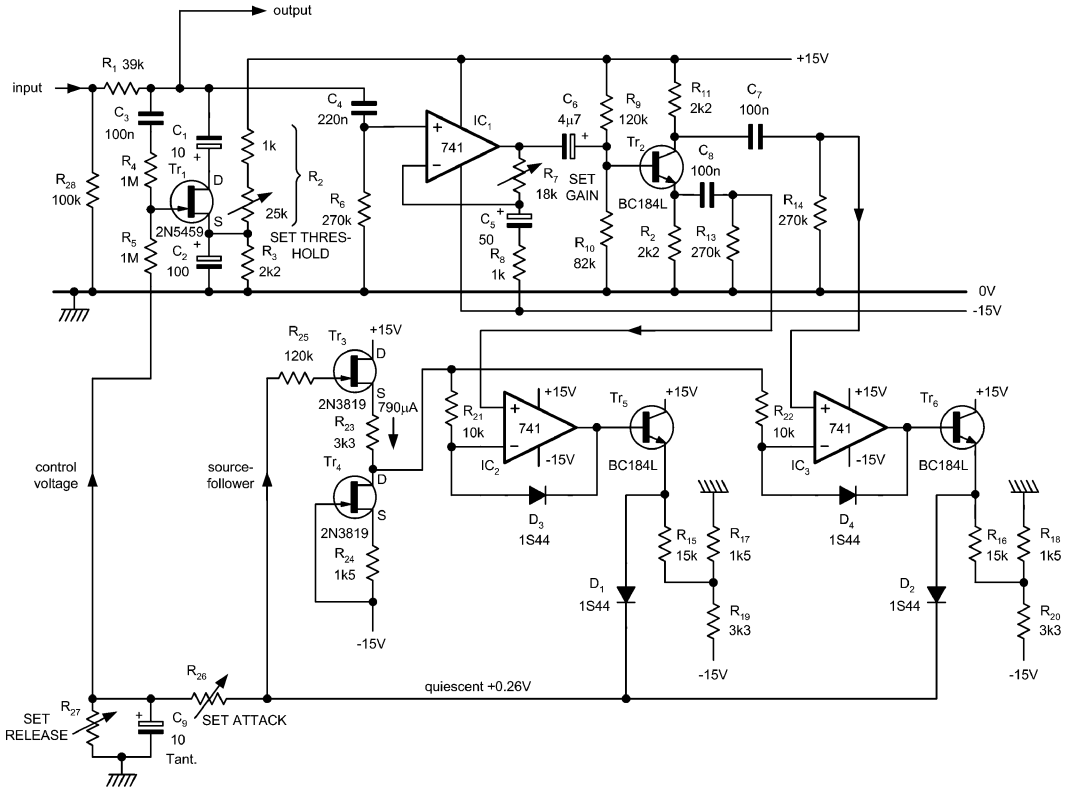



**Figure 6** *Complete circuit where the output is taken directly from the v.c.a. – this may be buffered for loads greater than* 100 kΩ.

for attenuation to begin is reduced, and the system's input/output characteristic moves smoothly from A to B on Figure 2. The threshold decreases and the compression slope becomes less flat as the system turns slowly from a limiter into a compressor by the manipulation of a single control. The output sensing amplifer consists of $IC_1$ and has a gain of 19 over the audio band. This is rolled off to unity at d.c. by $C_5$. Transistor $Tr_2$ and its associated components form a conventional phase-splitter driving $IC_2$ and $IC_3$ the precision rectifiers. The rectifier circuitry is more complex than implied above, three modifications have been made to improve the performance. Firstly, $IC_2$ and $IC_3$ charge $C_9$ via current amplifier stages $Tr_5$ and $Tr_6$ otherwise the current-limited 741 outputs would be unable to provide enough current for the faster attack times (less than 1 mS). Secondly, the feedback loop from $C_9$ to the inverting unputs of $IC_2$ and $IC_3$ is completed via a FET source-follower. Without this, $C_9$ would be loaded by the two 741 inputs, and this would severely limit the maximum decay

times available. Incorporating the source-follower allows decay times of several minutes by using large resistance values for $R_{27}$. The conventional source-follower has a large negative offset voltage and is unusable in this application because due to their the rectifying action $IC_2$ and $IC_3$ are unable to provide a voltage on $C_9$ that is negative of ground. This would be required to allow the source-follower output to be at ground when there is no input to the rectifers. However, if a modified source-follower is used, with a constant-current source and resistance combination in the source circuit, the offset voltage can be varied on either side of zero by manipulation of $R_{24}$ which varies the driving current. The offset voltage is arranged to be plus 0.3 V, to allow a large safety margin for thermal variations, component ageing, etc. This means that under no-signal conditions $C_9$ takes up a standing quiescent voltage of plus 0.3 V. The effect of this is taken up in the calibration of $R_2$.

The third modification is the addition of $R_{21}$, $D_3$, and $R_{22}$, $D_4$. These two networks prevent $IC_2$ and $IC_3$ from saturating negatively, during negative half-cycles of their input voltage, by allowing local negative feedback through $D_3$ and $D_4$. This limits the negative excursion of the IC outputs to about 2 V. The prevention of saturation is necessary because the recovery time of the 741s causes the frequency response of the precision rectifier circuit to drop off at about 1 kHz. The addition of the anti-saturation networks provides a frequency response that starts to fall off significantly above about 12 kHz which is ample for our purposes as program signals have very little energy content above this frequency.

The final part of the circuit defines the attenuation time constants. Resistor $R_{26}$ sets the attack time constant and $R_{27}$ the decay time constant; these can range between 0 and 1 MΩ (220 μs and 10 s) for $R_{26}$, and 1 kΩ and ∞ (10 mS and 20 min) for $R_{27}$. They can be either switched or variable resistances, depending on the range of variation required.

The circuit in Figure 6 shows the compressor output being taken directly from the v.c.a. This is only suitable if the minimum load to the output is greater than 100 kΩ, otherwise the v.c.a. attenuation characteristic will be distorted by excessive loading. If lower resistance loads are to be driven a buffer amplifier stage must be interposed. The $IC_1$ amplifier stage is suitable for most applications, and its gain is $(R_7 + R_8)/R_8$. For the unity gain case $R_8$ & $C_5$ can be eliminated and $R_7$ replaced by a direct connexion.

The compressor should be driven from a reasonably low impedance output (less than 5 kΩ).

Construction is straightforward; the layout is not critical and the prototype was assembled on 0.1 in matrix Veroboard. To set up the circuit $R_{24}$ is adjusted so that the voltage across $C_9$ is about +0.3 V with no signal input. The value required will vary due to production spreads in the f.e.ts. To calibrate $R_2$ it is necessary to relate the level of input signal at which attenuation commences, with the voltage across $C_2$. This can be

**Table 3** *Prototype calibration data and compression ratios*

| $VC_2$ (V) | *Threshold* (mV, pk) | *Compression ratio* |
|---|---|---|
| 2.9 | 10 | 2.3 |
| 3.5 | 20 | 5.1 |
| 5.0 | 50 | 10 |
| 6.7 | 100 | 20 |
| 8.5 | 200 | 35 |
| 9.8 | 500 | 50 |

done with an oscilloscope, or preferably an a.f. millivoltmeter. As a guide the calibration data for the prototype is shown in Table 3, along with the values of the compression ratio (number of dBs the input must increase by to increase the output by 1 dB). This data must be regarded as only a guide. It is worth noting that as the controlling factor setting the compression/limiting function is the voltage across $C_2$ $R_2$ could be replaced by a 1 kΩ resistor connected to a remote voltage source.

The compressor/limiter is quite straightforward in use, provided a few points are kept in mind. Firstly, if it is being used in the limiting mode to prevent overload of a subsequent device, the fastest possible attack time should be used, to catch fast transients, and a fast decay time (say 100 ms; $R_{27} = 10\,k\Omega$), to allow the system to recover rapidly when the transient has passed. Secondly, if a noisy programme signal is being compressed a long decay time should be employed, otherwise the noisy background will be faded up during quiet passages, and the familiar compressor 'breathing noises' will be heard. Finally, signals with a large v.l.f. content should be avoided or filtered, otherwise v.l.f. modulation of the signal will result, if a fast decay time is in use.

If a stereo compressor/limiter is constructed from two of the systems described above it is necessary to gang together $R_2$, $R_{26}$, and $R_{27}$ between the two channels. A direct connection between the non-grounded sides of the two $C_9$s is also needed. It might be necessary to select matched f.e.ts to avoid stereo image shift during compression, due to differing attenuation characteristics in the two v.c.as. A well-smoothed p.s.u. providing ±15 V should be used to power the compressor/limiter.

## Components list

| | |
|---|---|
| $IC_{1'2'3}$ | 741 |
| $Tr_{2'5'6}$ | BC184L or equivalent |

| | | |
|---|---|---|
| $Tr_1$ | 2N5459 | |
| $Tr_{3'4}$ | 2N3819 | |
| $D_{1'2'3'4}$ | IS44 or low-leakage equivalent | |
| $R_1$ | 39 k | |
| $R_2$ | 25 k variable, with 1 k in series | |
| $R_3$ | 2.2 k | |
| $R_{4'5}$ | 1 M | |
| $R_6$ | 270 k | |
| $R_7$ | 18 k | |
| $R_8$ | 1 k | |
| $R_9$ | 120 k | |
| $R_{10}$ | 82 k | All resistors |
| $R_{11'12}$ | 2.2 k | (except $R_2$) ¼ W |
| $R_{13'14}$ | 270 k | |
| $R_{15'16}$ | 15 k | |
| $R_{17'18}$ | 1.5 k | |
| $R_{19'20}$ | 3.3 k | |
| $R_{21'22}$ | 10 k | |
| $R_{23}$ | 3.3 k | |
| $R_{24}$ | see text | |
| $R_{25}$ | 120 k | |
| $R_{26'27}$ | see text | |
| $R_{28}$ | 100 k | |
| $C_1$ | 10 µF 25 V electrolytic | |
| $C_2$ | 100 µF 25 V electrolytic | |
| $C_3$ | 100 nF 250 V polyester | |
| $C_4$ | 220 nF 250 V polyester | |
| $C_5$ | 50 µF 40 V electrolytic | |
| $C_6$ | 4.7 µF 40 V electrolytic | |
| $C_{7'8}$ | 100 nF 250 V polyester | |
| $C_9$ | 10 µF 16 V tantalum bead | |

# 11 Inside mixers

*April 1991*

When this was written, mixing console design was my day job, so to speak, and I was doing hi-fi design purely as a spare-time pursuit. Having been responsible for many of the contemporary improvements in mixing console design (such as the padless mic amp and the active panpot) I thought it might be a good plan to publicise these in Electronics World. After consulting those set in authority over me, a fine balance was struck between offering solid, accurate technical content and not giving away too much to our competitors. The padless microphone amp and the active panpot could be described in some detail, as they were fully covered by patents.

I was a bit more cautious about the electronic switching, as that was not patented. The whole subject of electronic switching was eventually revealed in two articles in 2004 and they are also collected in this book.

Recording technology has changed greatly since this article was written, and the frequent references to tape-machines seem very dated. Yet, that was how it was then.

A large mixing console arguably represents the most demanding area of audio design. The steady advance of digital media demands that every part of the chain that takes music from performer to consumer must be near-perfect, as the comfortable certainty that everything will be squeezed through the quality bottleneck of either analogue tape or vinyl disc now looks very old-fashioned. This chapter was prompted by the introduction of the Soundcraft 3200 recording console, which is believed to have the highest performance in terms of noise, crosstalk and linearity of any console ever built.

Competition to sell studio time becomes more cut-throat with every passing week, and it is clear that advances in console quality must not harm cost-effectiveness. The only way to reconcile these demands is to innovate and to keep a very clear view as to what is really necessary to meet a demanding specification; in other words the way forward is to use

conventional parts in an unconventional way, rather than simply reaching for the most expensive op-amp in the catalogue.

The technical problems that must be over-come in a professional mixing console are many. A large number of signals flow in a small space and they must be kept strictly apart until the operator chooses to mix them; crosstalk must be exceedingly low.

There may be up to 64 input channels, each with many stages, and all having the potential to add distortion and noise to the precious signal. Even summing these signals together, while sounding trivially easy, is in practice a major challenge. In short, requirements are much more demanding than those for the most expensive hi-fi equipment, because degradation introduced at the recording stage can never be retrieved.

Major functions of consoles are largely standardised, although there is much scope for detailed variation. Figure 1 shows a typical system diagram for a split (separate groups) mixing console. The technique of multi-track recording is explained in the appendix at the end of this article.

Figure 2 shows a typical input channel for a mixing console. The input stage provides switchable balanced mic and line inputs; the mic input has an impedance of 1–2 kΩ, which provides appropriate loading for a 200 Ω mic capsule, while the line input has a bridging impedance of not less than 10 kΩ. This stage gives a wide range of gain control and is followed immediately by a high-pass filter (usually −3 dB at 100 Hz) to remove low-frequency disturbances.

The tone-control section (universally known in the audio business as 'EQ' or equalisation) typically includes one or more mid-band resonance controls as well as the usual shelving Baxandall-type high and low controls. Channel level is controlled by a linear fader and the panpot sets the stereo positioning, odd group numbers being treated as left, and even as right. The prefade-listen (PFL) switch routes the signal to the master module independently of all other controls; a logic bus signals the master module to switch the studio monitoring speakers from the normal stereo mix bus to the PFL bus, allowing any specific channel to be examined in isolation.

Figure 3 shows a typical group module and Figure 4 the basics of a master section; a manual source-select switch allows quality checking of the final stereo recording and two solid-state switches replace the stereo monitor signal with the PFL signal whenever a PFL switch anywhere on the console is pressed.

## Microphone inputs

The microphone preamplifier is a serious design challenge. It must provide from 0 to 70 dB of gain to amplify deafening drum-kits or discreet dulcimers, present an accurately balanced input to cancel noise pickup in

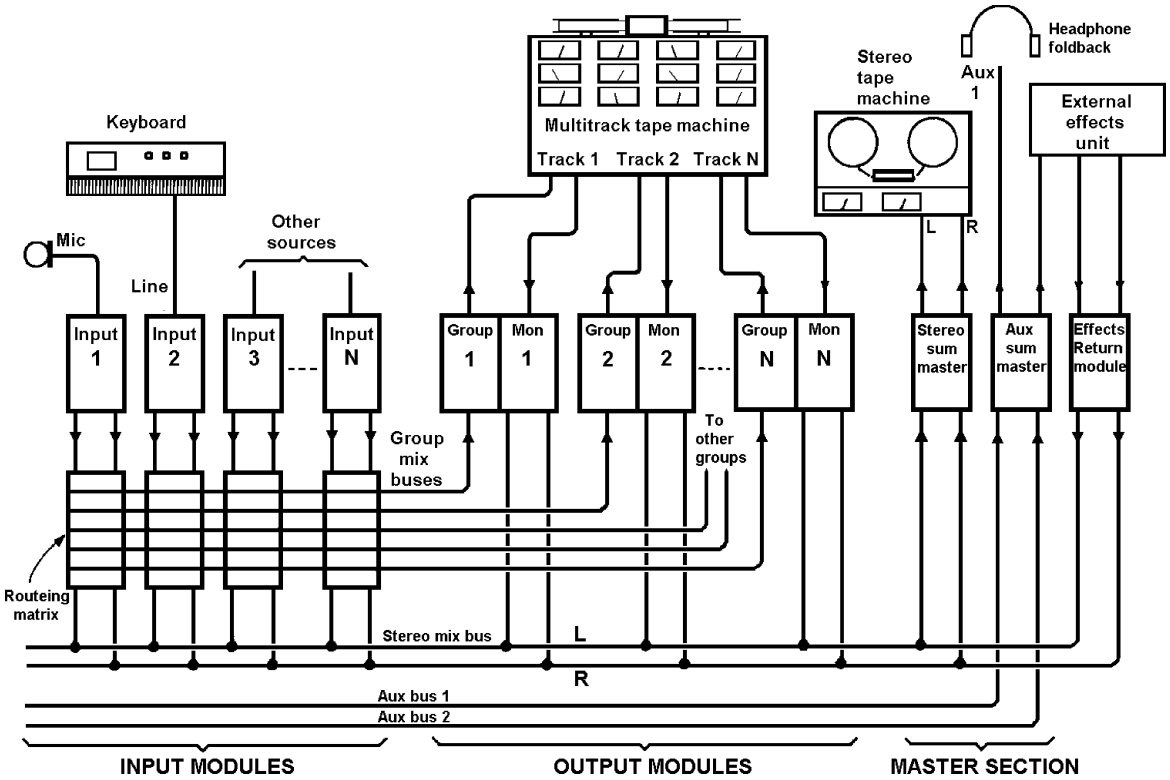



**Figure 1** *System diagram of complete mixing console, showing division into inputs, group monitor contributions and master modules. Routeing matrix determines which group of inputs shall be fed to a given track on the multi-track tape machine. Several channels share one effects device.*

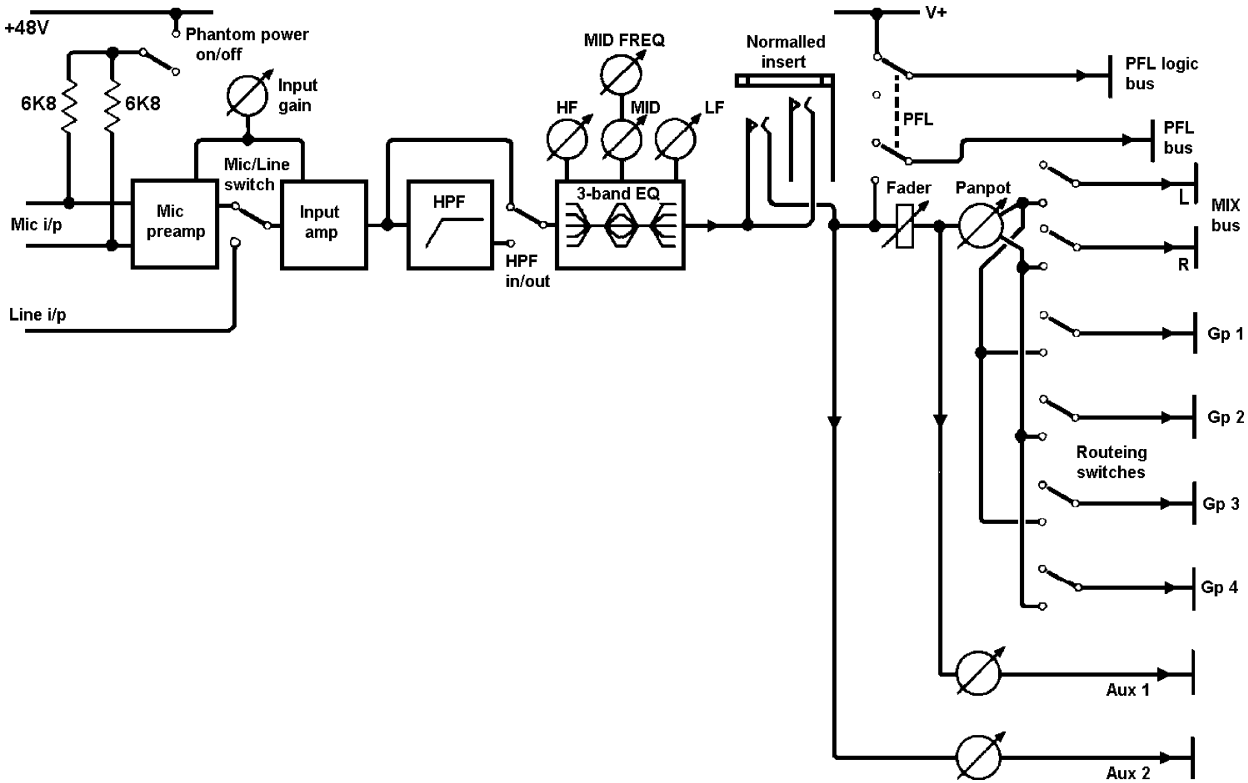



**Figure 2** *One input channel. Gain control is* 70 dB *and tone control is standard Baxandall shelving type with addition of mid-range lift and cut. Two auxiliary sends are shown.*

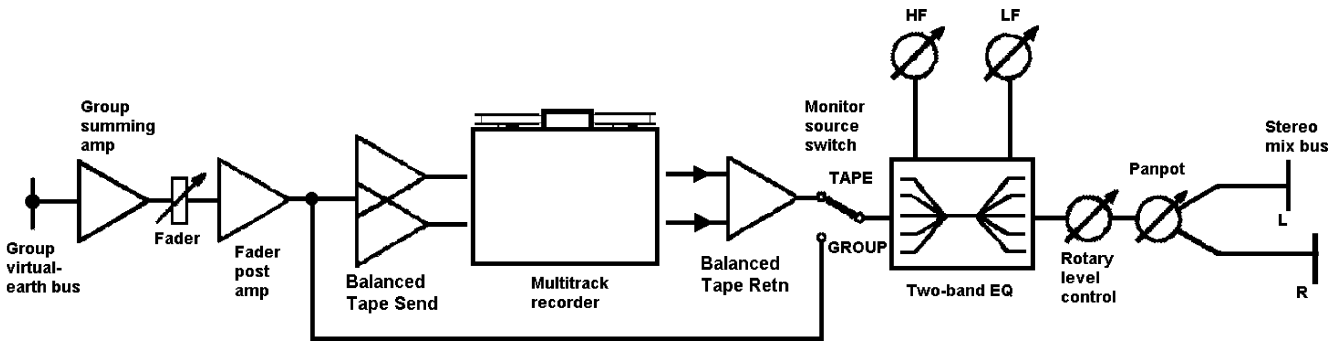



**Figure 3** *Block diagram of typical group module, showing switching between direct output and tape replay for monitoring purposes.*

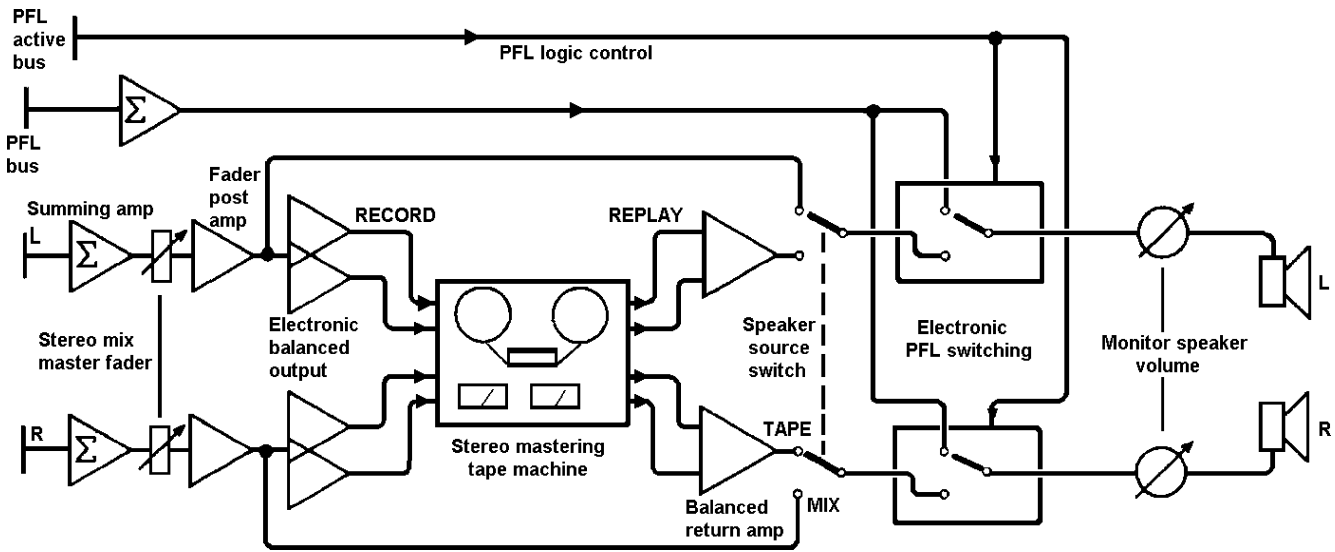



**Figure 4** *Block diagram of master module, with tape send/replay switching and automatic PFL switching.*

long cables and generate minimal internal noise. It must also be able to withstand +48 V DC suddenly applied to the inputs (for phantom-powering internal preamps in capacitor mics) while handling microvolt signals. The Soundcraft approach is to use standard parts, which are proven and cost-effective through quantity production, in new configurations. The latest mic preamplifier design, as used on the Series 3200, is new enough to be covered by patent protection.

It is now rare to use input transformers to match the low-impedance (150–200 Ω) microphone to the preamplifier, since the cost and weight penalty is serious, especially when linearity at low frequencies and high levels is important. The low-noise requirement rules out the direct use of op-amps, since their design involves compromises that make them at least 10 dB noisier than discrete transistors at low impedance.

This circuit, shown in Figure 5, therefore uses a balanced pair of low-noise, low-$R_b$ PNP transistors as an input stage, working with two op-amps to provide load-driving capability and raw open-loop gain to linearise signal handling. Preamplifier gain is spread over two stages to give a smooth 0–70 dB gain range with the rotation of a single knob. This eliminates the switched 20 dB attenuator that is normally required to give the lower gain values, not only saving cost and complication, but also avoiding the noise deterioration and CMRR degradation that switched attenuators impose. The result is an effective input stage that is not only quieter, but also more economical than one using specialised low-noise op-amps.

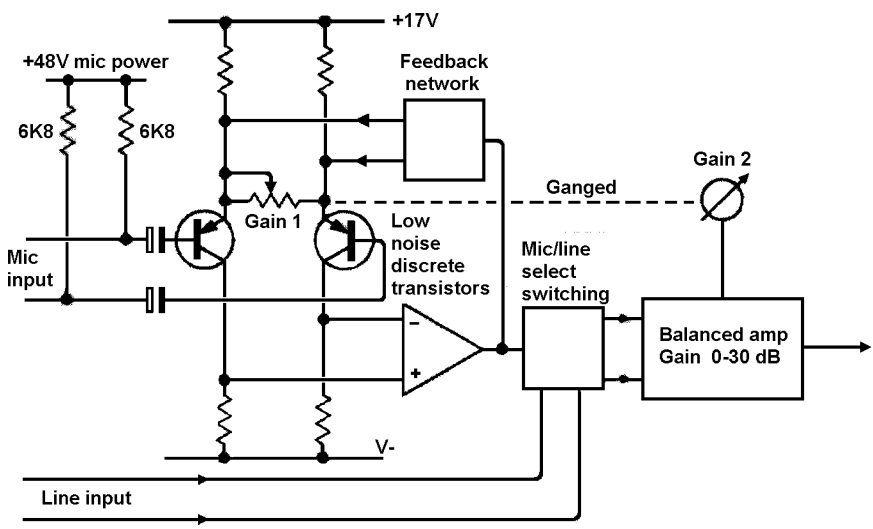



**Figure 5** *Low-noise microphone amplifier with wide gain range and balanced line output. Transistors in first stage avoid noise problem of op-amps.*

## Equalisation

Since large recording consoles need sophisticated and complex tone-control systems, unavoidably using large numbers of op-amps, there is a danger that the number of active elements required may degrade the noise performance.

A typical mid-band EQ that superimposes a +15 dB resonance on the flat unity-gain characteristic is shown in Figure 6. A signal is tapped from the forward path, put through a state-variable band-pass filter which allows control of centre-frequency and Q, and then added back. To improve noise performance, the signal level at all locations (in all conditions of frequency, Q, and boost/cut) was assessed, and it proved possible to double the signal level in the filter over the usual arrangement, while maintaining full headroom. The signal returned into the forward path is then attenuated to maintain the same boost/cut, and the noise added is thus reduced by about 6 dB.

## Auxiliary sends: foldback and effects

The auxiliary sends of a console represent an extra mixing system that works independently of the main groups; the number and configuration of these sends have a large effect in determining the overall versatility of

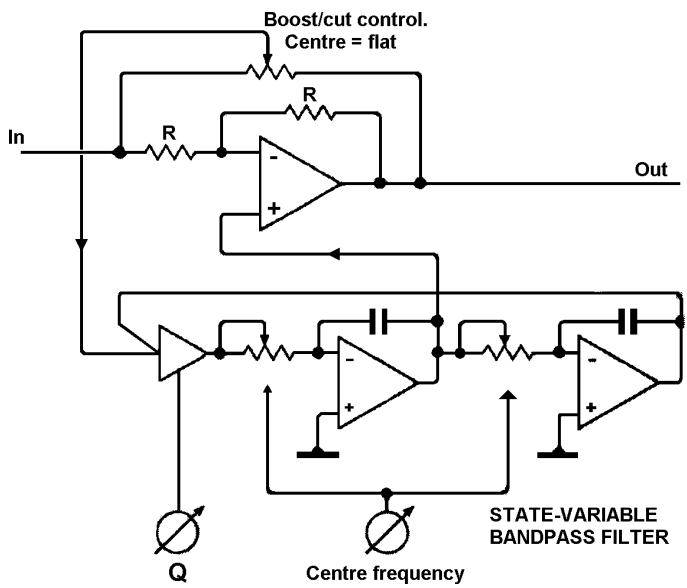



**Figure 6** *Parametric mid-band EQ stage. EQ and centre frequencies are independently variable, being set by the parameters of the state-variable filters.*

the console. Each send control provides a feed to a console-wide bus; this is centrally summed and then sent out of the console.

Sends come essentially in two kinds: prefade sends, which are taken from before the main channel fader, and post-fade sends, which take their feed from after the fader, so that the final level depends on the settings of both. There may be anything from one to twelve sends available, often switchable between pre and post. Traditionally, this means laboriously pressing a switch on every input module, since it is most unlikely that a mixture of pre and post sends on the same bus would be useful; the Series 3200 minimises the effort by setting pre/post selection for each bus from a master switch that controls solid-state pre/post switching in each module.

Prefade sends are normally used for 'foldback'; i.e. sending the artist a headphone feed of what he/she is perpetrating, which is important if electronic manipulation is part of the creative process, and essential if the artist is adding extra material that must be in time with that already recorded. In the latter case, the existing tracks are played back to the artist via the prefade sends on the monitor sections.

Postfade sends are used as effects sends; their source is after the fader, so that the effect will be faded down at the same rate as the untreated signal, maintaining the same ratio. The sum of all feeds to a given bus is sent to an external effects unit and the output of this returned to the console. This allows many channels to share one expensive device (this is particularly applicable to digital reverb) and is often more appropriate than the alternative of patching a processor into the channel insert point.

'Effect returns' may be either modules in their own right or a small subdivision of the master section. The returned effect, which may well now be in stereo, the output of a digital reverb, for example, is usually added to the stereo mix bus via level and pan controls. EQ is also sometimes provided.

## Panpot

To give smooth stereo panning without unwanted level changes, the panpot should theoretically have a sine/cosine characteristic; such components exist, but they are prohibitively expensive and so most mixing consoles use a dual linear pot. with its law bent by a pull-up resistor, as shown in Figure 7(a).

This not only gives a mediocre approximation of the required law, but also limits the panning range, since the pull-up signal passes through the wiper contact resistance (usually greater than the end-of-track resistance) and limits the attenuation the panpot can provide when set hard left or right. This limitation is removed in the Soundcraft active panpot shown

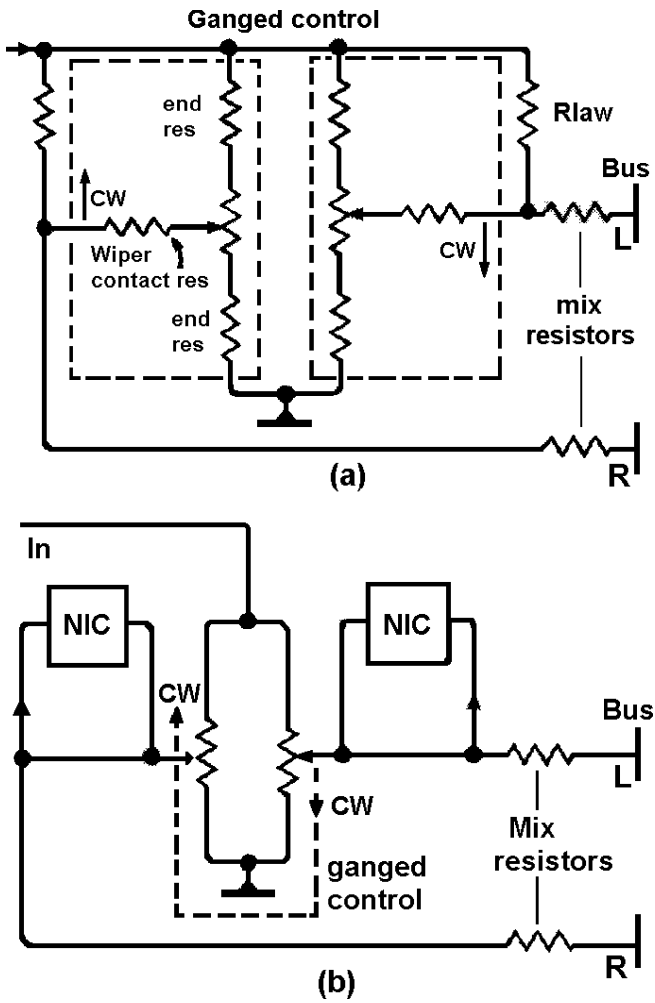



**Figure 7** *Standard panpot circuit at (a) showing how pull-up resistor draws current through wiper contact resistance, which is usually greater than the end resistance of the pot., limiting maximum attenuation. Arrangement at (b) uses NICs to replace pull-up to modulate law with panpot setting. Left/right isolation increased from* −65 dB *to* −90 dB.

in Figure 7b by replacing the pull-up with a negative-impedance-converter that modulates the law-bending effect in accordance with the panpot setting, making a close approach to the sine law possible. There is no pull-up at the lower end of the wiper travel, when it is not required, so the left-right isolation using a good-quality pot is improved from approx −65 to −90 dB. This has also been made the subject of patent protection.

## Summing

One of the main technical challenges in console design is the actual mixing of signals. This is done almost (but not quite) universally by virtual-earth techniques, as in Figure 8(a). A summing amplifier with shunt feedback is used to hold a long mixing bus at apparent ground, generating a sort of audio black hole; signals fed into this via mixing resistors apparently vanish, only to reappear at the output of the summing amplifier, as they have been summed in the form of current. The elegance of virtual-earth mixing, as opposed to the voltage-mode summing technique in Figure 8(b), is that signals cannot be fed back out of the bus to unwanted places, as it is effectively grounded, and this can save massive numbers of buffer amplifiers in the inputs.

There is, however, danger in assuming that a virtual earth is perfect; a typical opamp summer loses open-loop gain as frequency increases, making the inverting input null less effective. The 'bus residual' (i.e. the voltage measurable on the summing bus) therefore increases with frequency and can cause inter-bus crosstalk in the classic situation with adjacent buses running down an IDC cable.

Increasing the number of modules feeding the mix bus increases the noise gain; in other words the factor by which the noise of the summing amplifier is multiplied. In a large console, which might have 64 inputs, this can become distinctly problematic. The Soundcraft solution is to again exploit the low noise of discrete transistors coupled to fast opamps, in configurations similar to the mic preamps.

These sum amplifiers have a balanced architecture that inherently rejects supply-rail disturbances, which can otherwise affect LF crosstalk performance.

As a console grows larger, the mix bus system becomes more extensive, and therefore more liable to pick up internal capacitive crosstalk or external AC fields. The 3200 avoids internal crosstalk by the use of a proprietary routeing matrix construction which keeps the unwanted signal on a bus down to a barely measurable 120 dB. This is largely a matter of keeping signal voltages away from the sensitive virtual-earth buses. Further improvement is provided by the use of a relatively low value of summing resistor; this also keeps the noise down, although since it drops as the square-root of the resistor value, at best, there is a clear limit to how far this approach will work before drive power becomes excessive; 4.7 kΩ is a reasonable minimum value.

External magnetic fields, which are poorly screened by the average piece of sheet steel, are rejected by the balanced nature of the Series 3200 mix buses, shown in Figure 8c. The operation is much the same as a balanced input; each group has two buses, which run physically as close together as possible and the group reads the difference between the two, effectively

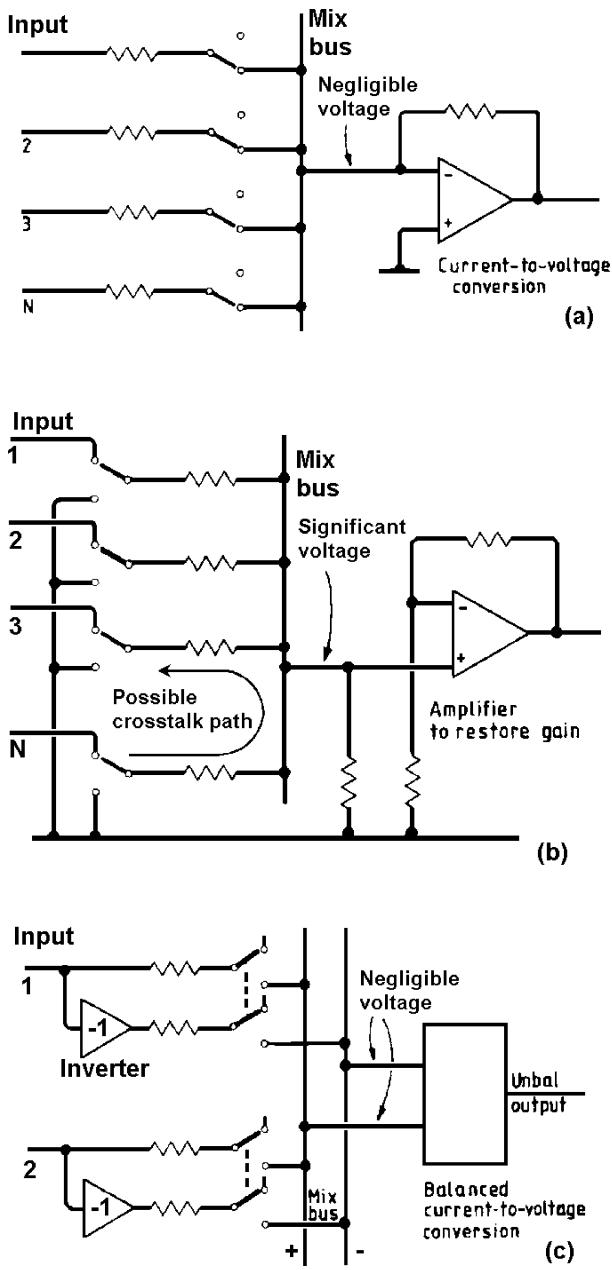



**Figure 8** *Virtual-earth summer at (a) effectively eliminates cross-talk, since there is almost no signal at the summing point. Voltage-mode circuit at (b) allows cross-talk. Balanced virtual-earth summing circuit at (c) requires a separate inverter for each channel to provide the antiphase signal.*

rejecting unwanted pickup. The two buses are fed in antiphase from each input, effectively doubling the signal level possible for a given supply voltage. Overall mixing noise is reduced by 3 dB, the signal level is 6 dB up and the noise, being uncorrelated for each bus, only increases by 3 dB.

The obvious method of implementing this is to use two summing amplifiers and then subtract the result. In the 3200, this approach is simplified by using one symmetrical summing amplifier to accept the two antiphase mix buses simultaneously; this reduces the noise level as well as minimising parts cost and power consumption. The configuration is very similar to that of the balanced mic amp, and therefore gives low noise as well as excellent symmetry.

## Solid-state switching

There are two main applications for electronic switching in console design. The first is 'hard' switching to reconfigure signal paths, essentially replacing relays with either JFETs (Figure 9(a)) or 4016-type analogue gates which, since they are limited to 18 V rails and cannot handle the full voltage swing of an opamp audio path, must be used in current mode, as shown in Figure 9(b). Note that when gate 1 is off, gate 2 must be on to ensure that a large voltage does not appear on gate 1 input. Full voltage range gates do exist but are very expensive.

Secondly, there is channel muting; this not a hard switch, since an unacceptable click would be generated unless the signal happened to be at a zero-crossing at the instant of switching; the odds are against you. The Series 3200 therefore implements muting as a fast-fade that takes about 10 ms; this softens transients into silence while preserving time-precision. It is implemented by a series-shunt JFET circuit, with carefully synchronised ramp voltages applied to the FET gates.

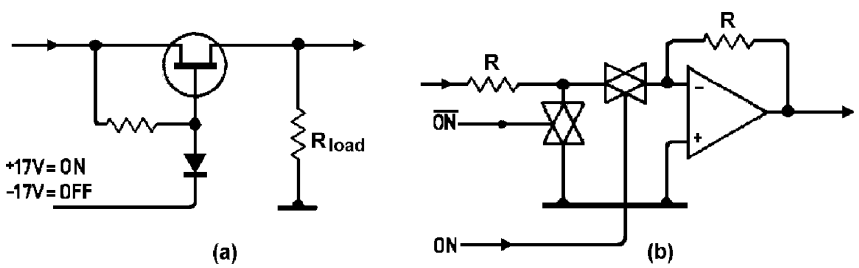



**Figure 9** *Hard switching with JFETs in voltage mode (a) and with analogue gates in the current mode (b), which prevents gate elements from being driven outside their voltage capabilities.*

## Performance factors

Primary requirements of modern consoles are very low noise and minimal distortion. Since a comprehensive console must pass the audio through a large number of circuit stages (perhaps over 100 from microphone to final mixdown) great attention to detail is essential at each stage to prevent a buildup of noise and distortion; the most important tradeoff is the impedance of the circuitry surrounding the opamp, for if this too high Johnson noise will be increased, while if it is too low an opamp will exhibit nonlinearity in struggling to drive it.

The choice of device is also critical, for cost considerations discourage the global use of expensive chips. In a comprehensive console like the 3200 with many stages of signal processing, this becomes a major concern; nonetheless, after suitable optimisation, the right-through THD remains below 0.004% at 20 dB above the normal operating level. At normal level it is unmeasurable.

## Appendix: the technique of multitrack recording

Multitrack recording greatly enhances the flexibility of recording music. The availability of a number of tape tracks (anywhere between 4 and 32 on one reel of tape) that can be recorded and played back separately allows each instrument a dedicated track, the beauty of this being that one mistake does not ruin the whole recording; only a single part need be done again. The multitrack process is in two basic halves; recording individual tracks (or 'tracklaying') and mixdown to stereo.

### *Recording*

Normally only one or two parts are recorded at once, though it quite possible to dedicate five or six tracks to a drum kit. The initial sound, whether captured by a microphone or fed in directly from a synthesiser line output, is usually processed as little as possible before committing it to tape; subsonic filtering and perhaps compression or limiting are used, but most effects are carefully avoided because they are usually impossible to undo later. You can easily add reverberation, for example, but just try removing it.

Recording is performed via the input modules, this being the only place where microphone preamps are fitted. The inputs are mixed together into groups if required; performers doing backing vocals might use four or five microphones, but these would almost certainly be mixed down to a stereo pair of groups at the recording stage, so that only two tape tracks are taken up. A bank of switches on each input module determines which group

shall be fed; this is known as the routing matrix. Combined group outputs are then sent to tape; however a 'group' is usually used even if only one signal is being recorded, as this is the part of the console permanently connected to the multitrack.

It is clearly essential that new parts are performed in time with the material already on tape and also that the recording engineer can make up a rough impression of the final mix as recording proceeds. Thus continually replaying already-recorded material is almost as important as recording it in the first place. During recording, the tape tracks already laid down are replayed through 'monitor sections' which are usually much-simplified inputs giving limited control; this keeps the more flexible inputs free for material that is actually being recorded. One of the major features of the Series 3200 is that the monitor sections are unusually capable, having facilities almost identical to the inputs and allowing much more accurate assessment of how the mix is progressing, reducing learning time for operators.

## *Mixdown*

When the tracklaying process is complete, there are 16 or more separate tape tracks that must be mixed down to stereo. Major manipulations of sound are done at this mixdown stage; since the multitrack tape remains unaltered, the resulting stereo being recorded on a separate two-track machine, any number of experiments can be performed without doing anything irrevocable.

Multitrack replay signals now enter the console through the input channels, so that the maximum number of facilities are available. Linear channel faders set the relative levels of the musical parts, while the rotary panpots (panoramic potentiometers) define the placement of instruments in the stereo sound field by setting the proportion of signal going to left and right mix buses. The monitor sections are now redundant, and can therefore be used either as extra inputs to the stereo mix, perhaps for keyboards, or to return effects.

## *Virtual mixing*

The advent of computer-based sequencers has given rise to the term 'virtual mixing'. Keyboard/synthesiser parts of the musical masterwork are not committed to multitrack, but instead stored in the form of MIDI sequencer data. This can be replayed at any time, providing means of synchronising it to the acoustic parts on the multitrack exist; this requires one tape track to be dedicated to some form of timecode.

The advantages are, firstly, that this gives almost any number of extra 'virtual tracks', and secondly, that the synthesiser parts suffer minimal degradation as they avoid one generation of tape storage.

# 12 Electronic analogue switching, Part I: CMOS gates

*January 2004*

In the course of the years spent designing mixing consoles, I came to know a great deal about analogue switching. All but the most basic mixers have a PFL system – the acronym meaning Pre-Fade Listen. In other words you can press a button on an input channel and have its contribution alone heard through your monitor speakers without disturbing the flow of signals to the main outputs. This requires the feed to the monitors to be switched, and for many years the only way to do this was a double-pole relay going 'clunk' somewhere in the console. It was a great relief to all concerned when the arrival of the 4016 analogue switch meant that this function could be performed electronically. It was cheap, reliable, and acoustically silent, and if the linearity was not perfect it was quite good enough. The chapter on mixing console design 'Inside Mixers' in this book gives more details on PFL systems and the like.

But why did discrete FET switching, as described in the second of these chapters, become so popular? It needs a little explaining. As mixing consoles grew more complex and sophisticated in the Seventies and Eighties, they came to incorporate much more electronic switching. The 'In-line' mixer format, in which input channel and output group lived in the same module and shared some facilities, became popular as analog tape machines grew from 8 to 16, to 24-track, because it gave a much more compact console. The downside of this was that there was only one equaliser (tone-control) section, which

needed to be used by the input channel while recording and the output group at mixdown time. It therefore had to be switchable from one signal path to the other, preferably by a single global control rather than pressing 48 or more individual switches. This demanded electronic switching – it takes six switches to move an equalisation section from one path to another and replace it with a direct connection in the path where it is not – as the cost, weight and power consumption of relays made an electromechanical solution unthinkable. However, the alternative of using 4016s was not much better – you will see from the first article that good linearity and the ability to handle a full op-amp output swing with these ICs requires the use of current-mode, with a virtual-earth (shunt-feedback) amplifier. This lands with you with an extra op-amp for each switch, which needs power and board space, and gives added noise and distortion in return. It also, crucially, introduces a phase inversion which must be immediately undone with *another* op-amp, because the preservation of correct phase in a mixing console is an absolute necessity. In contrast, discrete FET switching can implement a phase-preserving, very linear switch with one JFET, a diode and a resistor, and that is why this technology was developed, by me amongst others, in the early Eighties.

## Electronic switching

The switching and routing of analogue signals is a fundamental part of signal processing, but not one that is easily implemented if accuracy and precision are required. This chapter focuses on audio applications, but the basic parameters such as isolation and linearity are equally relevant in many fields.

Any electronic switching technique must face comparison with relays, which are still very much with us. Relays give total galvanic isolation between control and signal, zero contact distortion, and in audio terms have virtually unlimited signal-handling capability. They introduce negligible series resistance and shunt leakage to ground is usually not even worth thinking about. Signal offness can be very good, but as with other kinds of switching, this depends on intelligent usage. There will always be capacitance between open contacts, and if signal is allowed to crosstalk through this to nominally off circuitry, the 'offness' will be no better than other kinds of switching. (Throughout this chapter I use the word 'offness' – which is not found in any spellchecker but is widely used in the pro audio sector – as the quickest way of referring to the ratio in dB by which an unwanted input is suppressed.)

Obviously relays have their disadvantages. They are big, expensive, and not always as reliable as more than a hundred years of development should

have made them. Their operating power is significant. Some kinds of power relay can introduce disastrous distortion if used for switching audio because the signal passes through the magnetic soft-iron frame; however such problems are likely to be confined to the output circuits of large power amplifiers. For small-signal switching the linearity of relays can normally be regarded as perfect.

Electronic switching is usually implemented with CMOS analogue gates, of which the well-known 4016 is the most common example, and these are examined first. However, there are many special applications where discrete JFETs provide a better solution, so these are dealt in the second part.

## Part 1: analogue gates

CMOS analogue gates, also known as transmission gates, are quite different from CMOS logic gates, though the underlying process technology is the same. Analog gates are bilateral, which means that either of the in/out leads can be the input or output; this is most emphatically not true for logic gates. The 'analogue' part of the name emphasises that they are not restricted to any fixed logic levels, but pass through whatever signal they are given with low distortion. The 'low' word there requires a bit of qualification, as will be seen later.

There is no 'input' or 'output' marked on these gates, as they are symmetrical. When switched on, the connection between the two pins is a resistance which passes current in each direction as usual, depending on the voltage between the two gate terminals.

Analogue gates have been around for a long time, and are in some ways the obvious method of electronic switching. They do however have significant drawbacks.

Analogue gates such as the 4016 are made up of two MOS FETs of opposite polarity connected back to back. The internal structure of a 4016 analogue gate is shown in Figure 1. The two transmission FETs with their protective diodes are shown on the right; on the left is the control circuitry. A and B are standard CMOS inverters whose only function is to sharpen up the rather soggy voltage levels that 4000-series CMOS logic sometimes provides. The output of B directly controls one FET, and inverter C develops the anti-phase control voltage for the FET of opposite polarity, which naturally requires an inverted gate voltage to turn it on or off.

MOS FETS are of the enhancement type, requiring a voltage to be applied to the gate to turn them on; (in contrast JFETs work in depletion mode and require gate voltage to turn them off) so the closer the channel gets to the gate voltage, the more the device turns off. An analogue gate with only one polarity of FET would be of doubtful use because Ron would become very high at one extreme of the voltage range. This is

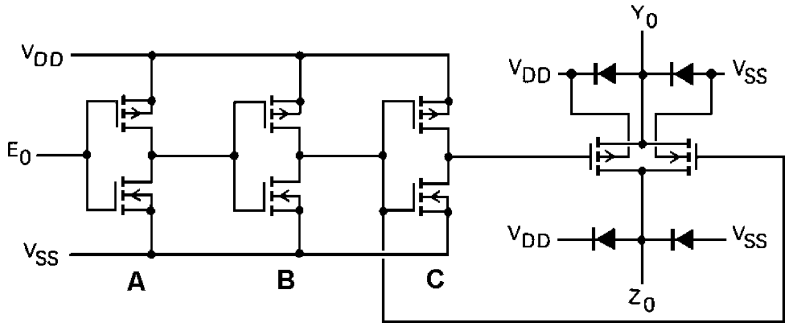



**Figure 1** *The internal circuitry of a* 4000*-series analogue gate.*

why complementary FETs are used; as one polarity finds its gate voltage decreasing, turning it off, the other polarity has its gate voltage increasing, turning it more on. It would be nice if this process cancelled out so the Ron was constant, but sadly it just doesn't work that way. Figure 2 shows how Ron varies with input voltage, and the peaky Ron curve gives a strong hint that something is turning on as something else turns off.

Figure 2 also shows that Ron is lower and varies less when the higher supply voltage is used; since these are enhancement FETs the on-resistance

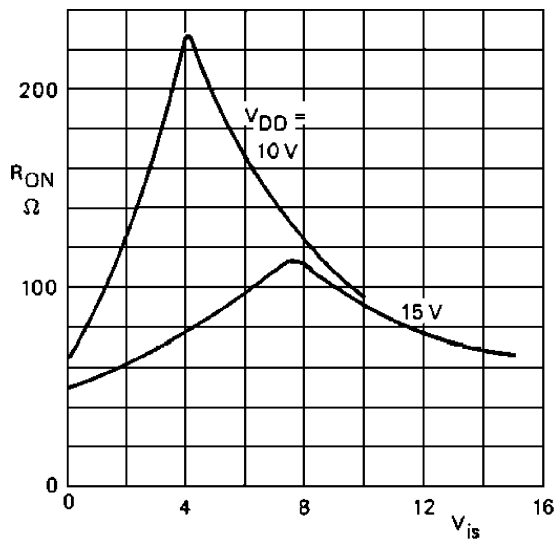



**Figure 2** *Typical variation of the gate series resistance Ron.*

decreases as the available control voltage increases. If you want the best linearity then always use the maximum rated supply voltage.

Since Ron is not very linear, the smaller its value the better. The 4016 Ron is specified as 115 Ω typical, 350 Ω max, over the range of input voltages and with a 15 V supply. The 4066 is a version of the 4016 with lower Ron, 60 Ω typical, 175 Ω max under the same conditions. This option can be very useful both in reducing distortion and improving offness, and in most cases there is no point in using the 4016. The performance figures given below assume the use of the 4066 except where stated.

## CMOS gates in voltage mode

Figure 3 shows the simplest and most obvious way of switching audio on and off with CMOS analog gates. This series configuration is in a sense the 'official' way of using them; the snag is that by itself it doesn't work very well.

Figure 4 shows the measured distortion performance of the simple series gate using the 4016 type. The distortion performance is a long way from brilliant, exceeding 0.1% just above 2 V r.m.s. These tests, like most in this section, display the results for a single sample of the semiconductor in question. Care has been taken to make these representative, but there will inevitably be some small variation in parameters like Ron. This may be greater when comparing the theoretically identical products of different manufacturers.

Replacing the 4016 gate with a 4066 gives a reliable improvement due to the lower Ron. THD at 2 V r.m.s. (10 K load) has dropped to a third of its previous level. There seems to be no downside to using 4066 gates instead of the more common and better-known 4016, and they are used exclusively from this point on.

The distortion is fairly pure second harmonic, except at the highest signal levels where higher-order harmonics begin to intrude. This is shown in Figures 5 and 6 by the straight line plots beginning to bend upwards above 2 V r.m.s.

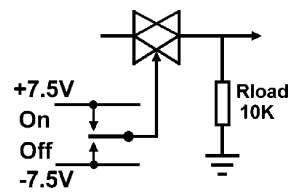



**Figure 3** *Voltage-mode series switching circuit using analogue gate.*

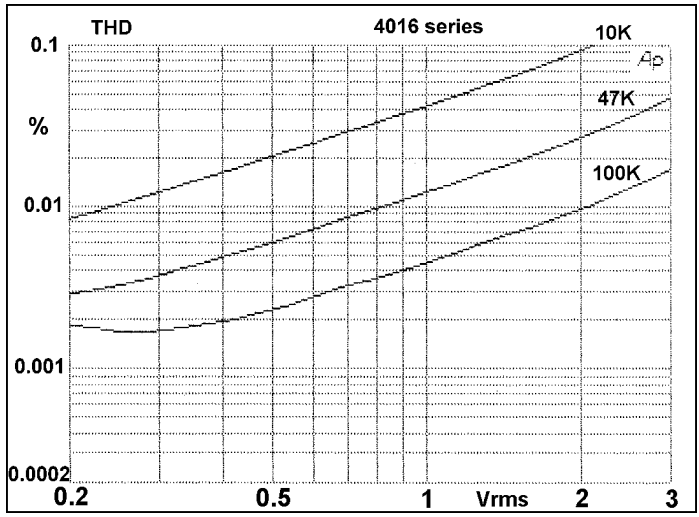



**Figure 4** *4016 series-gate THD versus level, with different load resistances.*

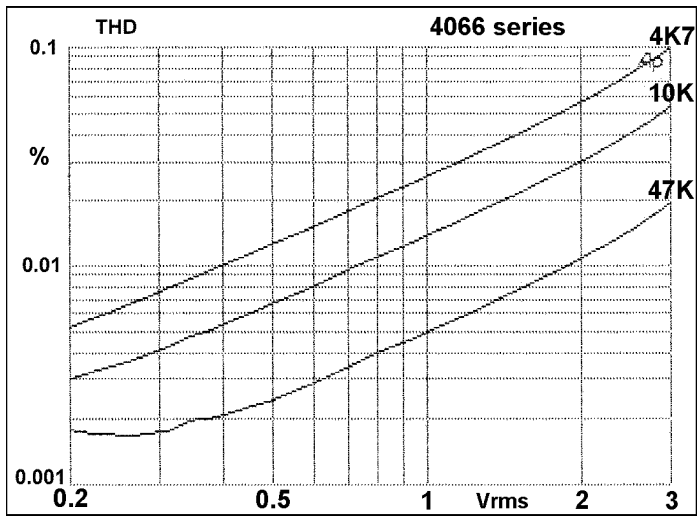



**Figure 5** *4066 THD versus level, with different load resistances.*

Analogue gate distortion is flat with frequency from 10 Hz up to 30 kHz at least, and so no plots of THD versus frequency are shown; they would merely be a rather uninteresting set of horizontal lines.

This circuit (Figure 3) gives poor offness when off, and poor distortion when on. The offness is limited by the stray capacitance in the package

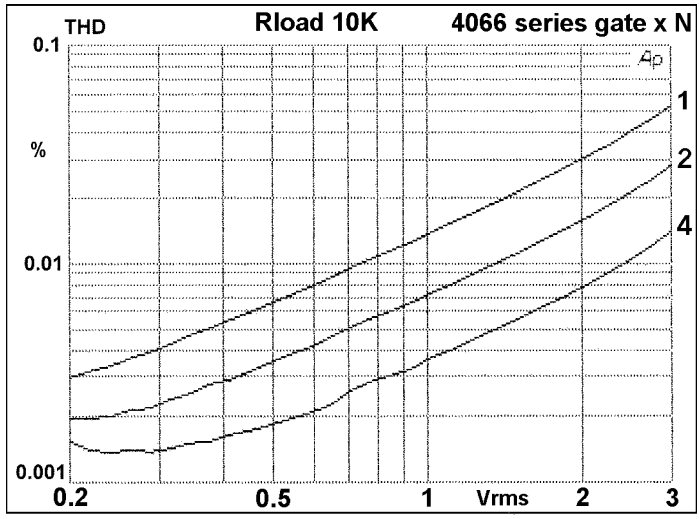



**Figure 6** *THD versus level, for different numbers of series* 4066 *gates.*

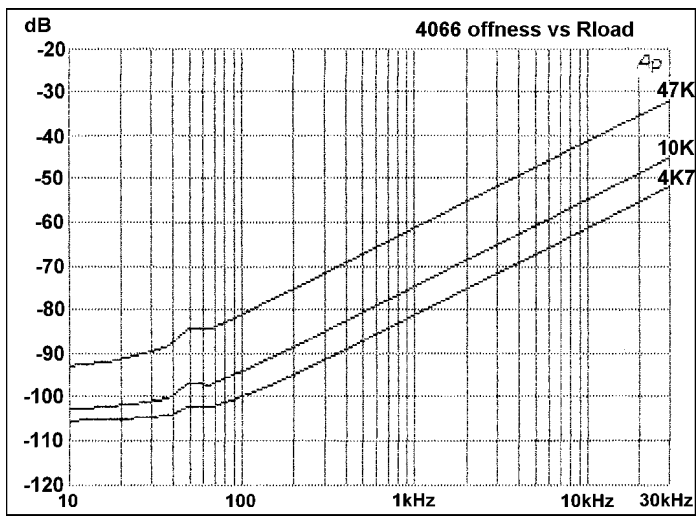



**Figure 7** *Offness versus load resistance.* −48 dB *at* 20 kHz *with a* 10 K *load.*

feeding through into the relatively high load impedance. If this is 10 K the offness is only −48 dB at 20 kHz, which would be quite inadequate for many applications. The load impedance could be reduced below 10 K to improve offness – for example, 4K7 offers about a 7 dB improvement – but this degrades the distortion, which is already poor at 0.055% for 3 V r.m.s.,

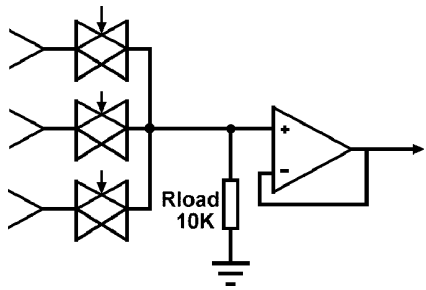



**Figure 8** *A one-pole, three way switch made from analogue gates.*

to 0.10%. Using 4066 gates instead of 4016 does not improve offness in this configuration. The internal capacitance that allows signals to leak past the gate seems to be the same for both types.

The maximum signal level that can be passed through (or stopped) is limited by the CMOS supply rails and conduction of the protection diodes. While it would in some cases be possible to contrive a bootstrapped supply to remove this limitation, it is probably not a good route to head down.

Figure 8 above shows a CMOS three-way switch. When analogue gates are used as a multi-way switch, the offness problem is much reduced, because capacitative feedthrough of the unwanted inputs is attenuated by the low Ron looking back into the (hopefully) low impedance of the active input, such as an opamp output. If this is not the case then the crosstalk from nominally off inputs can be serious. In this circuit the basic poor linearity is unchanged, but since the crosstalk problem is much less, there is often scope for increasing the load impedance to improve linearity. This makes Ron a smaller proportion of the total resistance. The control voltages must be managed so that only one gate is on at a time, if there is a possibility of connecting two opamp outputs together.

It may appear that if you are implementing a true changeover switch, which always has one input on, the resistor to ground is redundant, and just a cause of distortion. Omitting it is however very risky, because if all CMOS gates are off together even for an instant, there is no DC path to the opamp input and it will register its displeasure by snapping its output to one of the rails. This does not sound nice.

Figure 9 shows the offness of a changeover system, for two types of FET-input opamps. The offness is much improved to −87 dB at 20 kHz, an improvement of 40 dB over the simple series switch; at the high-frequency end however it still degrades at the same rate of 6 dB/octave. It is well-known that the output impedance of an op-amp with negative feedback increases with frequency at this rate, as the amount of internal gain falls, and this effect is an immediate suspect. However, there is actually no detectable signal on the opamp output, (as shown by the

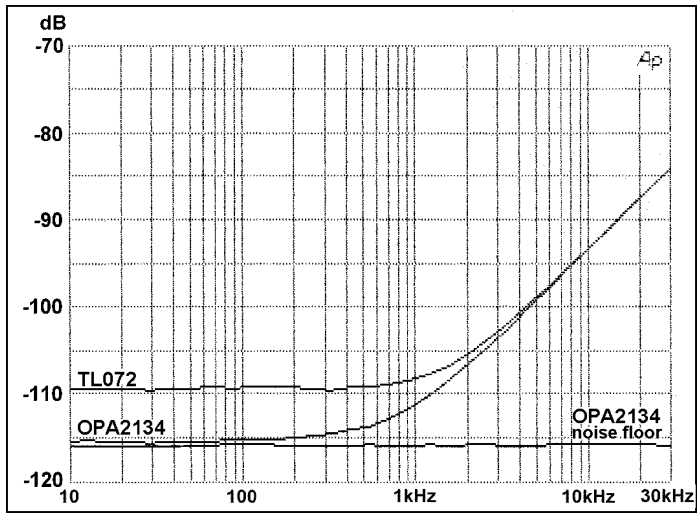



**Figure 9** *Voltage-mode changeover circuit offness for TL072 and OPA2134.* $R_{load} = 10\,K$.

lowest trace in Figure 9) and is also not very likely that two completely different opamps would have exactly the same output impedance. I was prepared for a subtle effect, but the true explanation is that the falling offness is simply due to feedthrough via the internal capacitance of the analogue gate.

It now remains to explain why the OPA2134 apparently gives better offness in the flat low-frequency region. In fact it does not; the flat parts of the trace represent the noise floor for that particular opamp. The OPA2134 is a more sophisticated and quieter device than the TL072, and this is reflected in the lower noise floor.

There are two linearity problems. Firstly, the on-resistance itself is not totally linear. Second, and more serious, the on-resistance is modulated when the gates move up and down with respect to their fixed gate voltages.

It will by now probably have occurred to most readers that an on/off switch with good offness can be made by making a changeover switch with one input grounded. This is quite true, but since much better distortion performance can be obtained by using the same approach in current mode, as explained below, I am not considering it further here.

Figure 10 above shows a shunt muting circuit. This gives no distortion in the 'ON' state because the signal is no longer going through the Ron of a gate. However the offness is limited by the Ron, forming a potential divider with the series resistor R; the latter cannot be very high in value or the circuit noise will be degraded. There is however the advantage that the offness plot is completely flat with frequency. Note that the ON and OFF states of the control voltage are now inverted.

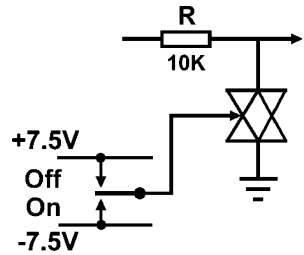



**Figure 10** *Voltage-mode shunt CMOS circuit.*

**Table 1** *Offness versus number of shunt gates*

| *N gates* | *Offness* (dB) |
|---|---|
| 1 | −37 |
| 2 | −43 |
| 4 | −49 |

Table 1 below gives the measured results for the circuit, using the 4066. The offness can be improved by putting two or more of these gates in parallel, but since doubling the number N only gives 6 dB improvement, it is rarely useful to press this approach beyond four gates.

## CMOS gates in current mode

Using these gates in current mode – usually by defining the current through the gate with an input resistor and dropping it into the virtual-earth input of a shunt-feedback amplifier – gives much superior linearity. It removes the modulation of channel resistance as the gate goes up and down with respect to its supply rails, and, in its more sophisticated forms, can also remove the signal voltage limit and improve offness.

Figure 11 shows the simplest version of a current-mode on/off switch, and it had better be said at once that it is a bit too simple to be very useful as it stands. An important design decision is the value of Rin and Rnfb, which are often equal to give unity gain. Too low a value increases the effect of the non-linear Ron, while too high a value degrades offness, as it makes the gate stray capacitance more significant, and also increases Johnson noise from the resistors. In most cases 22 K is a good compromise.

Table 2 gives the distortion for dBu (7.75 V r.m.s.) in/out, and shows that it is now very low compared with voltage-mode switchers working at much lower signal levels; compare with Figures 5 and 6. The increase in THD at

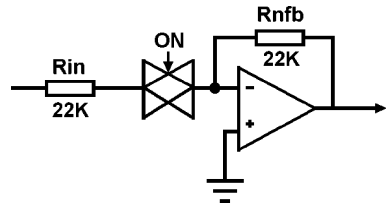



**Figure 11** *The simplest version of a current-mode on/off switch.*

**Table 2** *Distortion and offness for* 4016 *series current-mode switching*

| | 1 kHz | 10 kHz | 20 kHz |
|---|---|---|---|
| THD via 4016, dBu (%) | 0.0025 | 0.0039 | 0.0048 |
| THD: 4016 shorted, dBu (%) | 0.0020 | 0.0036 | 0.0047 |
| Offness (dB) | −68 | −48 | −42 |

high frequencies is due to a contribution from the opamp. However, the offness is pretty poor, and would not be acceptable for most applications. The problem is that with the gate off, the full signal voltage appears at the gate input and crosstalks to the summing node through the package's internal capacitance. In practical double-sided PCB layouts the inter-track capacitance can usually be kept very low by suitable layout, but the internal capacitance is inescapable.

In Figures 11 and 12, the CMOS gate is powered from a maximum of +/−7.5 V. This means that in Figure 11, signal breakthrough begins at an input of 5.1 V r.m.s. This is much too low for opamps running off their normal rail voltages, and several dB of headroom is lost. Figure 12 shows a partial cure for this. Resistor Rin2 is added to attenuate the input signal when the CMOS gate is off, preventing breakthrough. There is no effect on gain when the gate is on, but the presence of Rin2 does increase the noise gain of the stage.

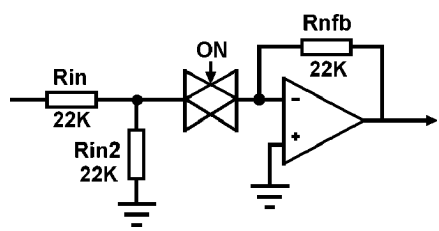



**Figure 12** *Current-mode switch circuit with breakthrough prevention resistor Rin2.*

## Series-shunt current mode

We now extravagantly use two 4016 CMOS gates, as shown in Figure 13.

When the switch is on, the series gate passes the signal through as before; the shunt gate is off and has no effect. When the switch is off the series gate is off and the shunt gate is on, sending almost all the signal at A to ground so that the remaining voltage is very small. The exact value depends on the 4016 specimen and its Ron value, but is about 42 dB below the input voltage. This deals with the offness (by greatly reducing the signal that can crosstalk through the internal capacitance) and also increases the headroom by several dB, as there is now effectively no voltage signal to breakthrough when it exceeds the rails of the series gate.

Two antiphase control signals are now required. An excellent way to generate the inverted control signal is to use a spare analogue gate as an inverter, as shown in Figure 14.

The distortion generated by this circuit can be usefully reduced by using two gates in parallel for the series switching, as in Table 3 below; this gate-doubling reduces the ratio of the variable Ron to the fixed series resistor and so improves the linearity. Using two in parallel is sufficient to

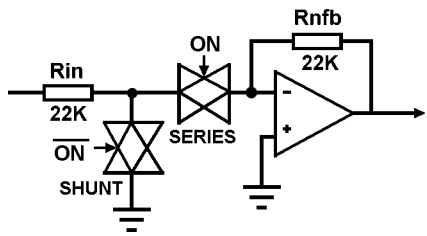



**Figure 13** *A series-shunt current-mode switch.*

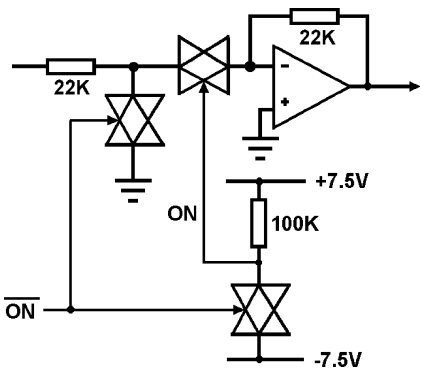



**Figure 14** *Generating the control signals with a spare analogue gate.*

**Table 3** *Distortion and offness for* 4016 *series-shunt current-mode switching*

| | 1 kHz | 10 kHz | 20 kHz |
|---|---|---|---|
| THD via 4016 × 1, +20 dBu | 0.0016% | 0.0026% | 0.0035% |
| THD via 4016 × 2, +20 dBu | 0.0013% | 0.0021% | 0.0034% |
| THD 4016 shorted, dBu | 0.0013% | 0.0021% | 0.0034% |
| Offness 4016 × 1 (dB) | −109 | −91 | −86 |
| Offness 4016 × 1, J111 (dB) | Less than −116 | −108 | −102 |

render the distortion negligible. (The higher distortion figures at 10 and 20 kHz are due to distortion generated by the TL072 opamp used in the measurements.)

As before the input and output levels are +20 dBu, well above the nominal signal levels expected in opamp circuitry; measurements taken at more realistic levels would show only noise.

Discrete FETs have lower Ron than analogue gates. If a J111 JFET is used as the shunt switching element the residual signal at A is further reduced, to about 60 dB below the input level, with a consequent improvement in offness, demonstrated by the final entry in Table 3. This could also be accomplished by using two or more CMOS gates in parallel for the shunt switching.

There is more on discrete FETs in Part Two of this article (Chapter 13).

## Control voltage feedthrough in CMOS gates

When an analogue gate changes state, some energy from the control voltage passes into the audio path via the gate-channel capacitance of the switching FETs, through internal package capacitances, and through any stray capacitance designed into the PCB. Since the control voltages of analogue gates move snappily, due the internal inverters, this typically puts a click rather than a thump into the audio. Attempts to slow down the control voltage going into the chip with RC networks are not likely to be successful for this reason. In any case, slowing down the control voltage change simply converts a click to a thump; the FET gates are moving through the same voltage range, and the feedthrough capacitance has not altered, so the same amount of electric charge has been transferred to the audio path – it just gets there more slowly.

The only certain way to reduce the effect of transient feedthrough is to soak it up in a lower value of load resistor. The same electric charge is applied to a lower resistor value (the feedthrough capacitance is tiny, and controls the circuit impedance) so a lower voltage appears. Unfortunately reducing the load tends to increase the distortion, as we have already seen; the question is if this is acceptable in the intended application.

# 13 Electronic analogue switching, Part II: discrete FETs

*February 2002*

## Discrete FET switching

Having looked in detail at analogue switching using CMOS gates, and having seen how well they can be made to work, you might be puzzled as to why anyone should wish to perform the same function with discrete FETs. There are at least two advantages in particular applications. Firstly, JFETs can handle the full output range of opamps working from maximum supply rails, so higher signal levels can often be switched directly without requiring opamps to convert between current and voltage mode. Secondly, the direct access to the device gate allows relatively slow changes in attenuation (though still measured in milliseconds, for reasons that will emerge) rather than the rapid on–off action which CMOS gates give as a result of their internal control-voltage circuitry. This is vital in creating mute circuits that essentially implement a fast fade rather than a sharp cut, and so do not generate clicks and thumps by abruptly interrupting the signal.

The downside is that they require carefully-tailored voltages to drive the gates, and these cannot always be conveniently derived from the usual opamp supply rails.

## Discrete FETs in voltage mode: the series JFET switch

The basic JFET series switching circuit is shown in Figure 1. With the switch open there is no other connection to the gate other than the bootstrap resistor, Vgs is zero, and so the FET is on. When the switch is closed, the gate is pulled down to a sufficiently negative voltage to ensure that the FET is biased off even when the input signal is at its negative limit.

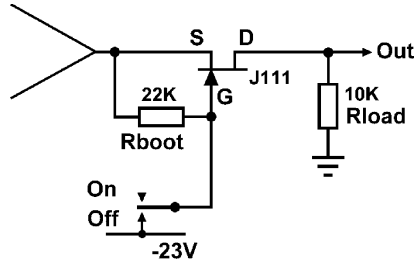



**Figure 1** *The basic JFET switching circuit, with gate bootstrap resistor.*

The JFET types J111 and J112 are specially designed for analogue switching and are pre-eminent for this application. The channel on-resistances are low and relatively linear. This is a depletion-mode FET, which requires a negative gate voltage to actively turn it off. The J111 requires a more negative Vgs to ensure it is off, but in return gives a lower Rds(on) which means lower distortion.

The J111, J112 (and J113) are members of the same family – in fact they are same the device, selected for gate/channel characteristics, unless I am much mistaken. Table 1 shows how the J111 may need 10 V to turn it off, but gives a 30 Ω on-resistance or Rds(on) with zero gate voltage. In contrast the J112 needs only 5.0 V at most to turn it off, but has a higher Rds(on) of 50 Ohms. The trade-off is between ease of generating the gate control voltages, and linearity. The higher the Rds(on), the higher the distortion, as this is a non-linear resistance.

FET tolerances are notoriously wide, and nothing varies more than the Vgs characteristic. It is essential to take the full range into account when designing the control circuitry.

Both the J111 and J112 are widely used for audio switching. The J111 has the advantage of the lowest distortion, but the J112 can be driven directly from 4000 series logic running from 7.5 V rails, which is often convenient. The J113 appears to have no advantage to set against its high Rds(on) and is rarely used – I have never even seen one.

The circuits below use either J111 or J112, as appropriate. The typical version used is shown, along with typical values for associated components.

**Table 1** *Characteristic of the J111 FET series*

| | *J111* | *J112* | *J113* |
|---|---|---|---|
| Vgs(off) min | −3.0 | −1.0 | −0.5 V |
| Vgs(off) max | −10 | −5.0 | −3.0 V |
| Rds(on) | 30 | 50 | 100 |

Figure 1 has Source and Drain marked on the JFET. In fact these devices appear to be perfectly symmetrical, and it seems to make no difference which way round they are connected, so further diagrams omit this. As JFETs, in practical use they are not particularly static-sensitive.

The off voltage must be sufficiently negative to ensure that Vgs never becomes low enough to turn the JFET on. Since a J111 may require a Vgs of −10 V to turn it off, the off voltage must be 10 V below the negative saturation point of the driving opamp – hence the −23 V rail. This is not exactly a convenient voltage, but the rail does not need to supply much current and the extra cost in something like a mixing console is relatively small.

To turn a JFET on, the Vgs must be held at 0 V. That sounds simple enough, but it is actually the more difficult of the two states. Since the source is moving up and down with the signal, the gate must move up and down in exactly the same way to keep Vgs at zero. This is done by bootstrap resistor Rboot in Figure 1. When the JFET is off, d.c. flows through this resistor from the source; it is therefore essential that this path be d.c.-coupled and fed from a low impedance such as an opamp output, as shown in these diagrams. The relatively small d.c. current drawn from the opamp causes no problems.

Figure 2 is a more practical circuit using a driver transistor to control the JFET. (If you had a switch contact available, you would presumably use it to control the audio directly.) The pull-up resistor Rc keeps diode D reverse-biased when the JFET is on; this is its sole function, so the value is not critical. It is usually high to reduce power consumption. I have used anything between 47 K and 680 K with success.

Sometimes d.c.-blocking is necessary if the opamp output is not at a d.c. level of 0 V. In this case the circuit of Figure 3 is very useful; the audio path is d.c.-blocked but not the bootstrap resistor, which must always have a d.c. path to the opamp output. Rdrain keeps the capacitor voltage at zero when the JFET is held off.

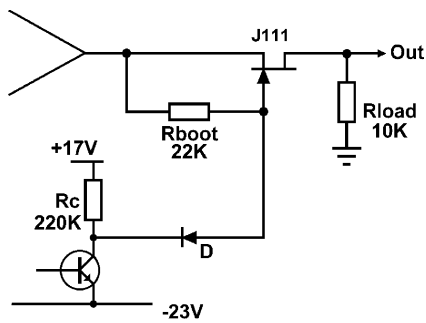



**Figure 2** *Using a transistor and diode for gate control.*

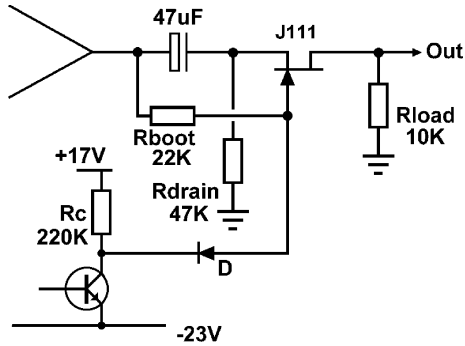



**Figure 3** *The JFET switching circuit with a d.c. blocking capacitor.*

Figure 4 shows the distortion performance with a load of 10 K. The lower curve is the distortion from the opamp alone; the low THD level should tell you immediately it was a 5532. The signal level was 7.75 V r.m.s. (+20 dBu).

Figure 5 shows the distortion performance with heavier loading, from 10 K down to 1 K. As is usual in the world of electronics, heavier loading makes things worse. In this case, it is because the non-linear Ron becomes a more significant part of the total circuit resistance. The signal level was 7.75 V r.m.s. (+20 dBu).

Figure 6 shows the distortion performance with different values of bootstrap resistor. The lower the value, the more accurately the drain follows the source at high audio frequencies, and so the lower the distortion. The signal level was 7.75 V r.m.s. (+20 dBu) once again. There appears to be no disadvantage to using a bootstrap resistor of 22 K or so, except in in special circumstances, as explained below.

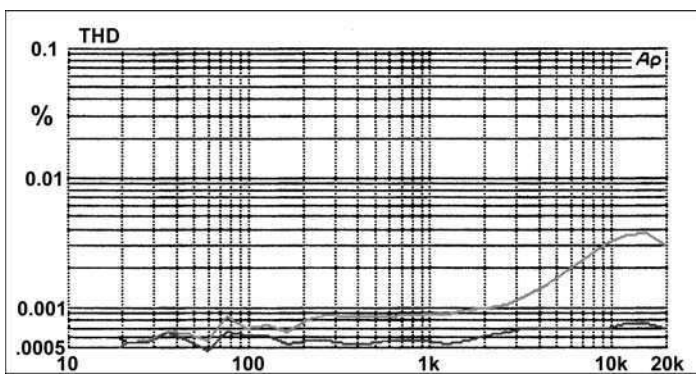



**Figure 4** *The JFET distortion performance with a load of* 10 K.

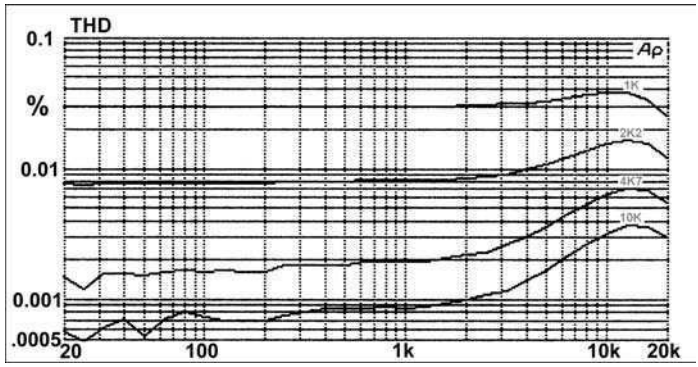



**Figure 5** *The JFET distortion performance versus loading.*

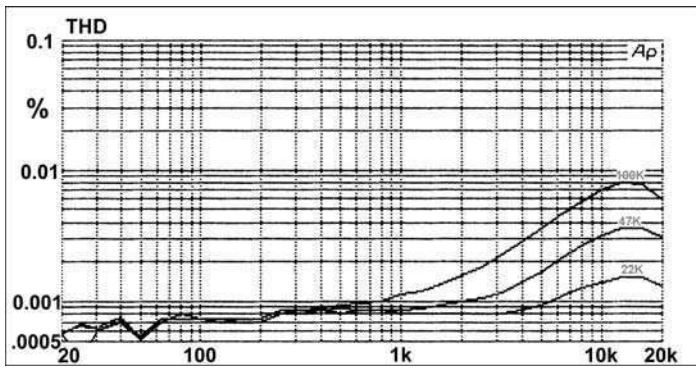



**Figure 6** *The distortion performance with different values of bootstrap resistor.*

Two series JFET switches can be simply combined to make a changeover switch, as shown in Figure 7. The valid states are A on, B on, or both off. Both on is not a good option because the two opamps will then be driving each other's outputs through the JFETs.

It is possible to cascade FET switches, as in Figure 8, which is taken from a real application. Here the main output is switched between A and B as before, but a second auxiliary output is switched between this selection and another input C by JFET3 and JFET 4. Cascading FET switches in this way removes the need for a buffer opamp between JFET1 and JFET3. The current drawn by the second bootstrap resistor Rboot2 must flow through the Rds(on) of the first FET, and will thus generate a small click. Rboot2 is therefore made as high as possible to minimise this effect, accepting that the distortion performance of the JFET3 switch will be compromised at HF; this was acceptable in the application as the second output was not a major signal path. The bootstrap resistor of JFET4 can be the desirable lower value of 22 K as this path is driven direct from an opamp.

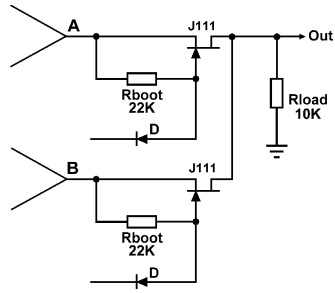



**Figure 7** *A JFET changeover switch.*

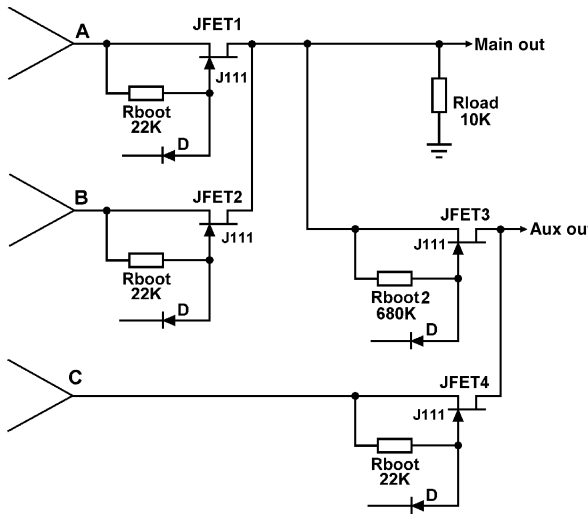



**Figure 8** *Cascaded FET switches.*

The offness of this type of series FET switch is not usually an issue because it is almost always used in the changeover format, where capacitative crosstalk from the off-JFET is made negligible by the low resistance of the on-JFET. If you simply want to turn a signal off, there are better ways to do it; see below.

## The shunt JFET switch

The basic JFET shunt switching circuit is shown in Figure 9. Like the shunt analogue gate mute, it gives poor offness but good linearity in the ON state, so long as its gate voltage is controlled so it never allows the JFET

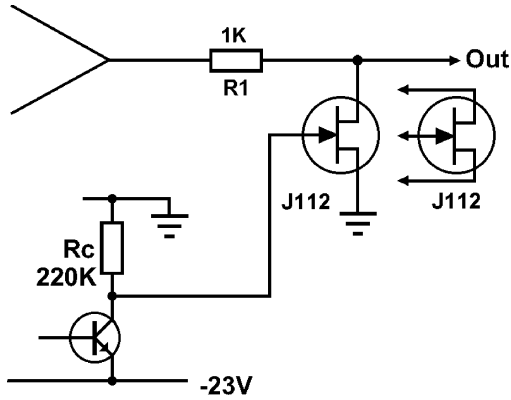



**Figure 9** *The basic JFET shunt switching circuit. Adding more JFETs in parallel increases the offness, but each* −6 dB *requires doubling the number.*

to begin conducting. Its great advantage is that the depletion JFET will be in its low-resistance before and during circuit power-up, and can be used to mute switch-on transients. Switch-off transients can also be effectively muted if the drive circuitry is configured to turn on the shunt FETs as soon as the mains disappears, and keep them on until the various supply rails have completely collapsed.

The circuit of Figure 9 was used to mute the turn-on and turn-off transients of a hifi preamplifier. Since this is an output that is likely to drive a reasonable length of cable, with its attendant capacitance, it is important to keep R1 as low as possible, to minimise the possibility of a drooping treble response. This means that the Rds(on) of the JFET puts a limit on the offness possible. The output series resistor R1 is normally in the range 47–100 Ω, when it has as its only job the isolation of the output opamp from cable capacitance. Here it has a value of 1 K, which is a distinct compromise. The muting obtained with 1 K was not quite enough so two J111s were used in parallel, giving a further −6 dB of attenuation, and yielding in total −33 dB across the audio band, which was sufficient to render the transients inaudible. The offness is not frequency dependent as the impedances are all low and so stray capacitance is irrelevant.

## Discrete FETs in current mode

JFETs can be used in the current mode, just as for analogue gates. Figure 10 shows the basic muting circuit, with series FET switching only. Ring prevents breakthrough. The stage as shown has less than unity gain; this can be corrected by increasing Rnfb.

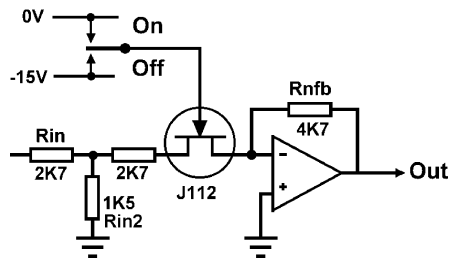



**Figure 10** *Circuit of series-only JFET mute bloc. Note phase-inversion. For performance see Figures 11 and 12.*

## *Soft JFET muting: crosstalk/linearity trade-off*

When switching audio signals, a instantaneous cut of the signal is sometimes not what is required. When a non-zero audio signal is abruptly interrupted there is bound to be a click. Perhaps surprisingly, clever schemes for making the instant of switching coincide with a zero-crossing give little improvement. There may no longer be a step-change in level, but there is still a step-change in slope and the ear once more interprets this discontinuity as a click.

What is really required is a fast-fade over about 10 ms. This is long enough to prevent clicks, without being so slow that the timing of the event becomes sloppy. This is normally only an issue in mixing consoles, where it is necessary for things to happen in real time. Such fast-fade circuits are often called 'mute blocks' to emphasise that they are more than just simple on–off switches. Analogue gates cannot be slowly turned on and off due to their internal circuitry for control-voltage generation. Therefore discrete JFETs must be used. Custom chips to perform this function have

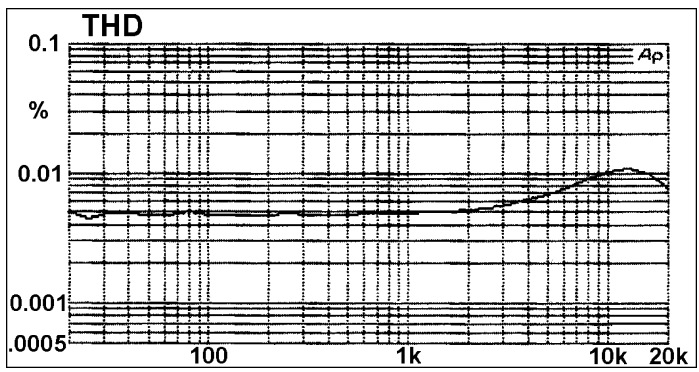



**Figure 11** *THD of the single-FET circuit in Figure 10 +20 dBu.*

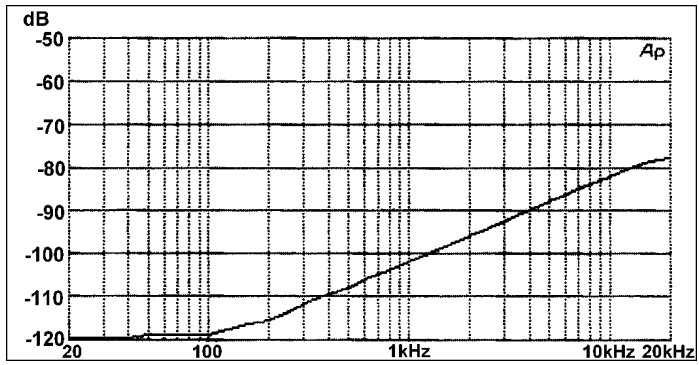



**Figure 12** *Offness of the single-FET circuit in Figure 10.*

been produced, but the ones I have evaluated have been expensive, single-source, and give less than startling results for linearity and offness. This situation is of course subject to change.

In designing a mute bloc, we want low distortion and good offness at the same time, so the series-shunt configuration, which proved highly effective with CMOS analogue gates, is the obvious choice. The basic circuit is shown in Figure 13. A small capacitor C is usually required to ensure HF stability, (Figure 14) due to the FET capacitances hanging on the virtual-earth node at D.

The control voltages to the series and shunt JFETs are complementary as before, but now they can be slowed down by RC networks to make the operation gradual, as shown in Figure 14. The exact way in which the control voltages overlap is easy to control, but the Vgs/resistance law of the FET is not (and is about the most variable FET parameter there is) and so the overlap of FET conduction is rather variable. However, I should say at once that this system does work, and works well enough to go in top-notch mixing consoles. The distortion performance is shown in Figure 15. As you go into the muted condition the series JFET turns off

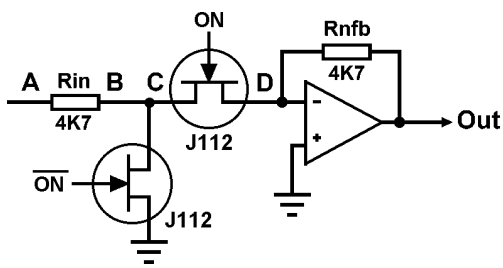



**Figure 13** *Series-shunt mode mute bloc circuit.*

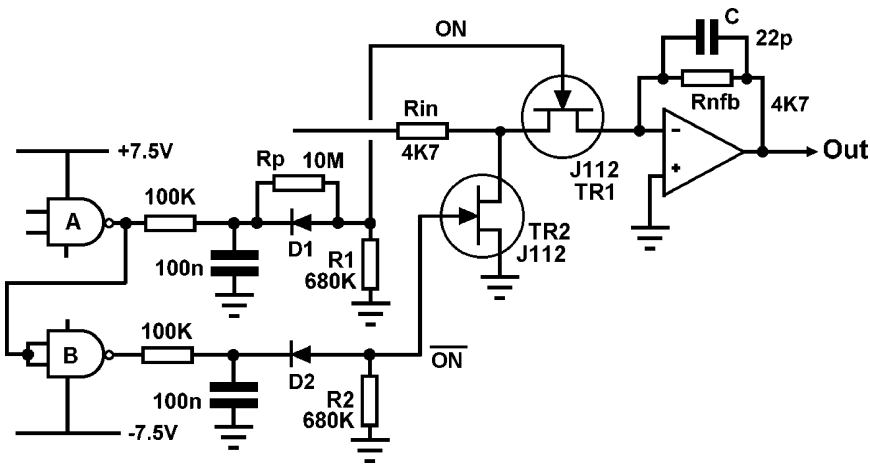



**Figure 14** *Circuitry to generate drive voltages for series-shunt JFET mute bloc.*

and the shunt JFET turns on, and if the overlap gets to be too much in error, the following bad things can happen:

1 If the shunt FET turns on too early, while the series JFET is still mostly on, a low-resistance path is established from the opamp virtual-earth point to ground, causing a large but brief rise in stage noise gain. This produces a 'chuff' of noise at the output as muting occurs.
2 If the shunt FET turns on too late, so the series JFET is mostly off, the large signal voltage presented to the series FET causes visibly serious distortion. I say 'visibly' because it is well-known that even quite severe distortion is not obtrusive if it occurs only briefly. The transition here is usually fast enough for this to be the case; it would however not be a practical way to generate a slow fade. The circuit of Figure 14 generates no audible distortion and only a very small chuff.

## *The drive circuitry*

The mute bloc requires two complementary drive voltages, and these are easily generated from 4000-series CMOS running from ±7.5 V rails. NAND gates are shown here as they are convenient for interfacing with other bits of control logic, but any standard CMOS output can be used. It is vital that the JFET gates get as close to 0 V as possible, ensuring that the series gate is fully on and gives minimum distortion, so the best technique is to and use diodes to clamp the gates to 0 V. Thus, in Figure 14, when the mute bloc is passing signal, the signal from NAND gate A is high, so D1 is reverse-biased

and the series JFET TR1 gate is held at 0 V by R1, keeping it on. (The role of Rp will be explained in a moment) Meanwhile, D2 is conducting as the NAND-gate output driving it is low, so the shunt JFET TR2 gate is at about −7 V and it is firmly switched off. This voltage is more than enough to turn off a J112, but cannot be guaranteed to turn off a J111, which may require −10 V (See Table 1). This is one reason why the J112 is more often used in this application – it is simpler to generate the control voltages. When the mute bloc is off, the conditions are reversed, with the output of A low, turning off TR1, and the output of B high, turning on TR2.

### *Reducing THD by on-biasing*

The distortion generated by this circuit bloc is of considerable importance, because if the rest of the audio path is made up of 5532 opamps – which is likely in professional equipment – then this stage can generate more distortion than the rest of the signal path combined, and dominate this aspect of the performance. It is therefore worth examining any way of increasing the linearity.

We have already noted that to minimise distortion, the series JFET should be turned on as fully as possible to minimise the value of the non-linear Rds(on). When a JFET has a zero gate-source voltage, it is normally considered fully on. It is, however, possible to turn it even more on than this.

The technique is to put a small positive voltage on the gate, say about 200–300 mV. This further reduces the Rds(on) in a smoothly continuous manner, without forward biasing the JFET gate junction and injecting d.c. into the signal path. This is accomplished in Figure 14 by the simple addition of Rp, which allows a small positive voltage to be set up across the 680 K resistor R1. The value of Rp is usually in the 10–22 MΩ range, for the circuit values shown here.

Care is needed with this technique, because if temperatures rise the JFET gate diode may begin to conduct after all, and d.c. will leak into the signal path, causing thumps and bangs. In my experience 300 mV is about the upper safe limit for equipment that gets reasonably warm internally, i.e. about 50 °C. Caution is the watchword here, for unwanted transients are much less tolerable than slightly increased distortion.

As with analogue CMOS gates, the choice of the resistors Rin and Rnfb that define the magnitude of the signal currents is an important matter. Figures 16 and 17 show the performance of the circuit is affected by using values of 4K7 and 22 K. Usually 4K7 would be the preferred value; choosing 22 K as the value makes the noise floor higher, as well as the signal leakage. Values below 4K7 are not usual as distortion will increase, as the JFET Rds(on) becomes a larger part of the total resistance in the circuit. The loading effect of Rin on the previous stage must also be considered.

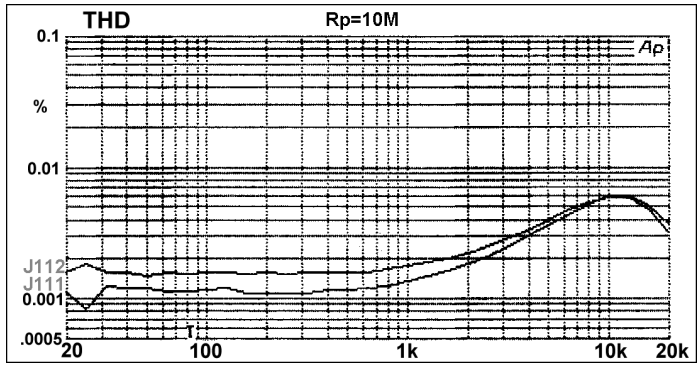



**Figure 15** *The THD of the mute bloc in Figure 13. The increase in FET distortion caused by using the J112 rather than J111 is shown +20 dBu.*

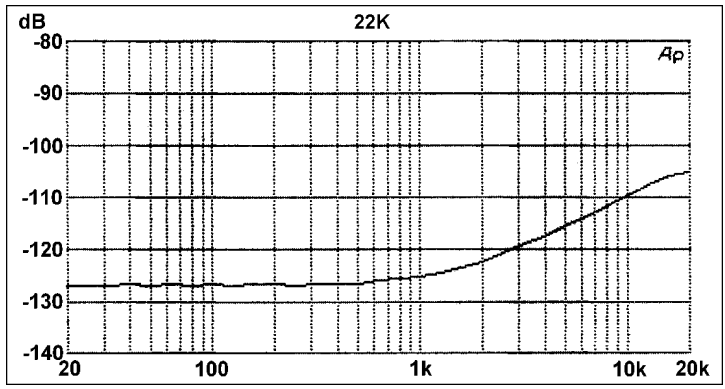



**Figure 16** *Offness of mute bloc in Figure 13 with Rin = Rnfb = 22 K.*

## *Layout and offness*

The offness of this circuit is extremely good, providing certain precautions are taken in the physical layout. In Figure 18, there are two possible crosstalk paths that can damage the offness. The path C–D, through the internal capacitances of the series JFET, is rendered innocuous as C is connected firmly to ground by the shunt JFET. However, point A is still alive with full amplitude signal, and it is the stray capacitance from A to D that defines the offness at high frequencies.

Given the finite size of Rin, it is often necessary to extend the PCB track B–C to get A far enough from D. This is no problem if done with caution. Remember that the track B–C is at virtual earth when the mute bloc is on, and so vulnerable to capacitative crosstalk from other signals straying into the area.

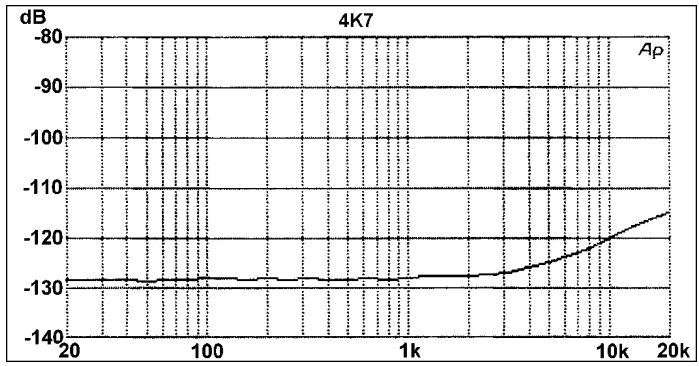



**Figure 17** *Offness of mute bloc in Figure 13 with Rin = Rnfb =* 4 K7*. Offness is 10* dB *better at 10* kHz *and the noise floor (the flat section below* 2 kHz*) has been lowered by about* 2 dB.

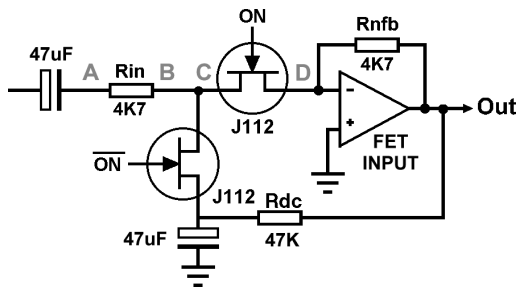



**Figure 18** *Circuit of JFET mute showing stray capacitances and d.c. handling.*

## *Dealing with the d.c.*

The circuits shown so far have been stripped down to their bare essentials to get the basic principles across. In reality, things are (surprise) a little more complicated. Opamps have non-zero offset and bias voltages and currents, and if not handled properly these will lead to thumps and bangs. There are several issues:

1 If there is any d.c. voltage at all passed on from the previous stage, this will be interrupted along with the signal, causing a click or thump. The foolproof answer is of course a d.c. blocking capacitor, but if you are aiming to remove all capacitors from the signal path, you may have a

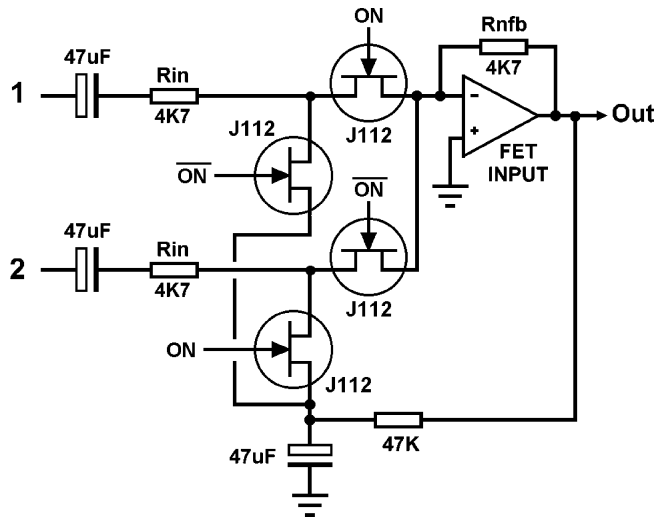



**Figure 19** *Circuit of JFET soft changeover switching.*

problem. d.c. servos can partly make up the lack, but since they are based on opamp integrators they are no more accurate than the opamp, while d.c. blocking is totally effective.

2 The offset voltage of the opamp. If the noise gain is changed when the mute operates (which it is) the changing amplification of this offset will change the d.c. level at the output. The answer is shown in Figure 18. The shunt FET is connected to ground via a blocking capacitor to prevent DC gain changes. This capacitor does not count as 'being in the signal path' as audio only goes through it when the circuit is muted. Feedback of the opamp offset voltage through $Rd_c$ to this capacitor renders it innocuous.

3 The input bias and offset currents of the opamp. These are potentially much more of a problem and are best dealt with by using JFET opamps such as the OPA2134, where the bias and offset currents are negligible at normal equipment temperatures.

## Soft changeover circuit

This circuit (Figure 19) is designed to give a soft changeover between two inputs – in effect a fast crossfade. It is the same mute block but with two separate inputs, either or both of which can be switched on. The performance at +20 dBu in/out is summarised in Table 2.

**Table 2** *Performance of the series-shunt JFET mute circuit in Figure 14*

| | 1 kHz | 10 kHz | 20 kHz |
|---|---|---|---|
| THD +20 dBu (%) | 0.0023 | 0.0027 | 0.0039 |
| Offness (dB) | −114 | −109 | −105 |

The THD increase at 20 kHz is due to the use of a TL072 as the opamp. J112 JFETs are used in all positions.

This circuit is intended for soft-switching applications where the transition between states is fast enough for a burst of increased distortion to go unnoticed. It is not suitable for generating slow crossfades in applications like disco mixers, as the exact crossfade law is not very predictable.

## Control voltage feedthrough in JFETs

All discrete FETs have a small capacitance between the gate and the device channel, so changes in the gate voltage will therefore cause a charge to be transferred to the audio path, just as for CMOS analogue gates. As before, slowing down the control voltage change tends to give a thump rather than a click to a thump; the same amount of electric charge has been transferred to the audio path, but more slowly. Lowering the circuit impedance reduces the effects of feedthrough, but can only be taken so far before distortion begins to increase as the non-linear Rds(on) of the JFET becomes a greater part of the total circuit resistance.

# 14 Sound mosfet design

*September 1990*

This was the first article on power amplifiers that I wrote, though I had been designing them for manufacture since 1975. This investigation into the concept of combining power FET output devices with bipolar drivers was done some years before I undertook my major investigation into the root causes of power amplifier distortion. (See Distortion in Power Amplifiers, Parts 1–8, later in this book.) In the throes of the design process, I realised with greater force than hitherto that the distortions in the small-signal part of a power amplifier were (a) far from negligible and (b) susceptible to analysis by a mixture of SPICE simulation and a few well-chosen experiments. I also determined that SPICE could be extremely useful in the analysis of output stages. The rest is history.

I have resisted the temptation to edit the text in the light of later knowledge, as this would, I think, be cheating. To prevent people being misled, I will just point out that the final circuit falls somewhat short of the Blameless performance standards set by the Distortion in Power Amplifiers series. The input stage may look symmetrical, but in fact is grotesquely unbalanced, with 16 μA flowing through the left transistor of the pair, and 580 μA through the right. This stage must have generated far more second-harmonic distortion than necessary, and with the benefit of hindsight I am not very proud of it. This imbalance also means that the input stage transconductance is far too low, which precludes any emitter degeneration of the input pair, and this also explains why a dominant-pole capacitor as small as 15 pF is enough for stability. The voltage-amplifier stage could be improved in linearity by adding another transistor within the $C_{dom}$ local feedback loop.

Mosfet amplifiers undoubtedly present a tantalising prospect for simple circuitry. Unfortunately, practical application requires careful consideration of their many foibles. But my aim here is to stimulate thinking about possible improvements to FET power amps, and describe two new avenues of development. Each produces a practical result, though neither should be regarded as a foolproof recipe for success.

Power-amplifier design is as prone as any other branch of audio to folklore and confusion. One of the less extreme myths holds that high levels of negative feedback are 'A Bad Thing' because they require heavier compensation for HF stability, leading to low slew rates and generally indolent and sluggish behaviour.

As far as it goes this is true. But only a poor designer would lose all sight of slew rate while adjusting amplifier compensation. Despite much study of TIM, DIM, SID, 'internal overload', 'delayed feedback' and the rest, everything comes back to slew rate.

If an amplifier can reproduce a 20 kHz sine wave at full amplitude without excessive distortion (say under 0.1%) it can be regarded as blameless in respect of speed. Apply as much feedback as is decent, but always keep an eye on stability and slew rate.

## Design fundamentals

Ground rules for this design study require DC output coupling, one preset only (for quiescent current) and as simple a circuit as possible.

In the current audio market, almost any technological approach appears to be acceptable (an idiosyncratic hybrid with valves driving power fets is one recent design,[1]) with the possible exception of capacitor output coupling.

While problems can include capacitor distortion at LF (of the real and measurable kind[2]) and perhaps grounding difficulties, the overwhelming simplicity of this method still has its attractions; the almost unnatural reliability of the capacitor-coupled Quad 303 should be mentioned at this point.

But a designer must still prove that he knows what a differential pair is, and d.c. coupling has therefore been adopted. However, remember that proper offset protection is not a trivial problem, as the cost of a reliable output relay and d.c. detection circuitry can add 30% to amplifier electronics costs. A relay seems unavoidable, as my own experiences with kamikaze crowbars have been distinctly unhappy, though I would not claim this to be a definitive judgment.

So a simple unregulated power supply looks to be the best. I realise this brings me into head-to-head confrontation with Mr Linsley Hood,[3] so I shall quickly explain my preference.

Putting expensive power semiconductors in a high-current dual supply can easily double an amplifier's electronic-component cost and there is much more to go wrong.

Ensuring PSU HF stability can be difficult, and the PSU compensation required, threatening a steadily rising output impedance versus frequency, can lead to some awkward amplifier stability problems.

Finally, the unregulated supply can deliver more power on a transient basis – which is exactly what is required for audio.

The price to be paid for unregulated simplicity is the attention to be paid to the amplifier's supply-rail rejection. But since it is physically impossible for the voltage on large reservoir capacitors to change very quickly, this rejection need only be extremely good at low frequencies.

The excellent supply-rail rejection of IC op-amps – and a power amplifier is, after all, only a big op-amp – shows the problem to be distinctly soluble, although I admit that op-amp PSRR often differs markedly between the two rails, and usually declines above 1 kHz. This sort of difficulty can be simply solved in a power amplifier with a little RC decoupling.

Harmonic distortion should be kept as low as possible, but without spending significant money specifically on its reduction. THD in commercial equipment varies more widely than any other performance parameter, ranging from 0.003% to 1% at similar powers[4] with the most expensive units often giving the worst performance.

In marketing circles, there are clearly two routes to take: make the THD vanishingly small to show you know what you're doing; or make it poor and imply that this very practical parameter has been sacrificed in favour of some intangible and unmeasurable sonic benefit.

I have always gone for the former and so have concentrated on linearity as being the prime determinant of amplifier topology.

Distortion performance is not easy to specify completely, ideally requiring a spectrum analysis of every combination of level, frequency and load impedance. But this is not practical and so I have summarised it as THD plotted against frequency, into 8 Ω, at different levels where appropriate.

Pros and cons of mosfet output devices have been thoroughly ventilated,[5,6] but one point needs qualification. They have been praised for having a large crossover region between the two halves of a Class B stage, but my experiments show this to be a very dubious advantage.

Mosfet outputs, with or without the augmentations described below, may be uncritical of quiescent current setting, but this really means that nothing is exactly right. Bipolar stages, with sharper crossover regions, do at least make it obvious where to set the quiescent current, providing it is set by observing the distortion residual – and it certainly should be.

Mosfet distortion residuals typically present a rather gnarled appearance, with plenty of harmonics at least to the seventh. Though it runs counter

to conventional wisdom, in my experience complementary-pair bipolar residuals tend to be smoother.

Another vital point is that mosfet complementary-pairs are rare, not particularly complementary, and definitely more expensive than the profusion of strictly N-channel devices intended for switching.

## Determining performance

Mosfet power amps have suffered more than most branches of technology from 'application-note cloning' though some original designs have been published, ranging from the complex[7,8] to the very complex.[9] These are well worth ferreting out, though perhaps unattractive commercially.

The 'standard' mosfet amplifier circuit (Figure 1) differs from the equally standard bipolar-style circuit (Figure 2) mainly in possessing a sort of push–pull/current-mirror configuration in the voltage amplifier stage, probably intended to provide better charge/discharge of the mosfet input capacitances. This stage is sometimes called the pre-driver, but it is less confusing to call the first full-voltage-swing stage 'the voltage amplifier stage', or VAS.

The first question to be asked is what improvement is made by this push–pull arrangement (Figure 3). The linearity is not very different and the benefit at HF is not startling. As always in science, it pays to be skeptical.

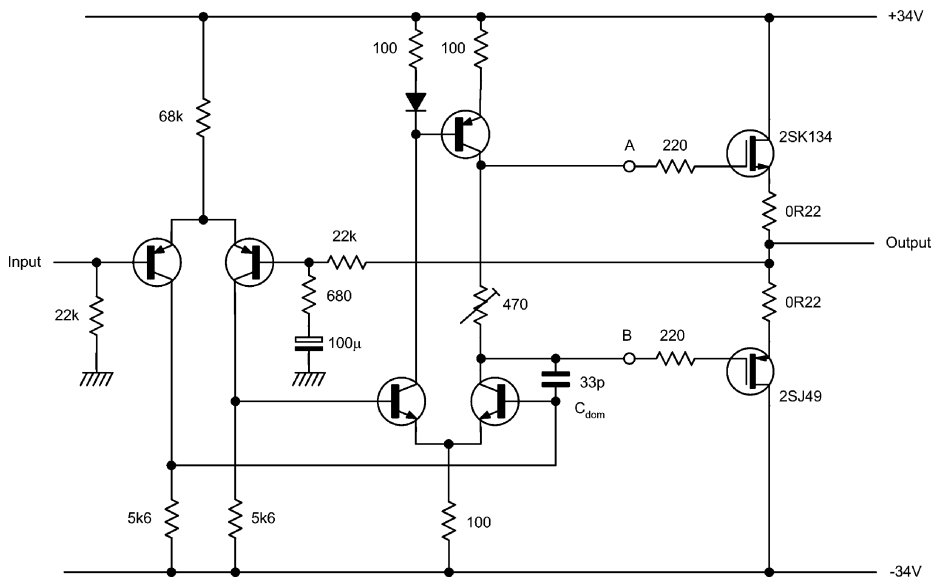



**Figure 1** *Standard mosfet power amplifier configuration.*

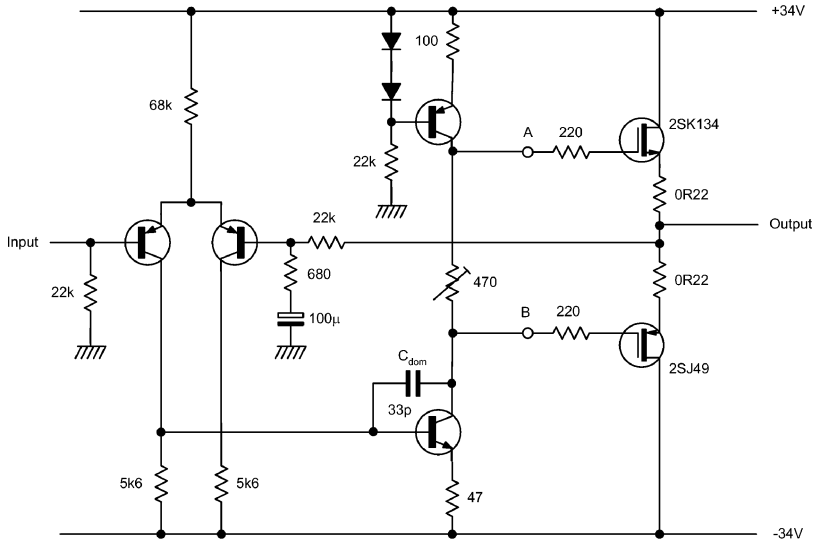



**Figure 2** *Standard bipolar power amplifier configuration, using mosfets.*

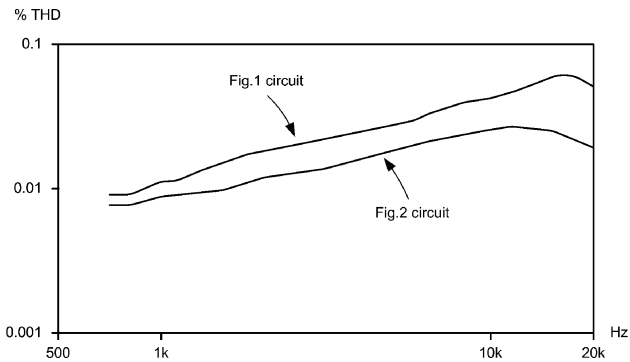



**Figure 3** *THD vs frequency for examples of Figures 1 and 2.* 10 V *r.m.s. into* 8 Ω.

Given that the simpler bipolar configuration of Figure 2 is workable, there are established ways to improve its overall linearity and some of these are shown in Figure 4.

Linearity can be enhanced simply by increasing open-loop gain (Figure 4a and b) or by a cascode arrangement (Figure 4c) which attempts to linearise the VAS by eliminating Early effect. Cascode arrangements are relatively ineffective at reducing distortion in power amplifiers, since the Early effect seems to be dominated by non-linear loading of the high impedance at the interface with the output stage (points A and B).

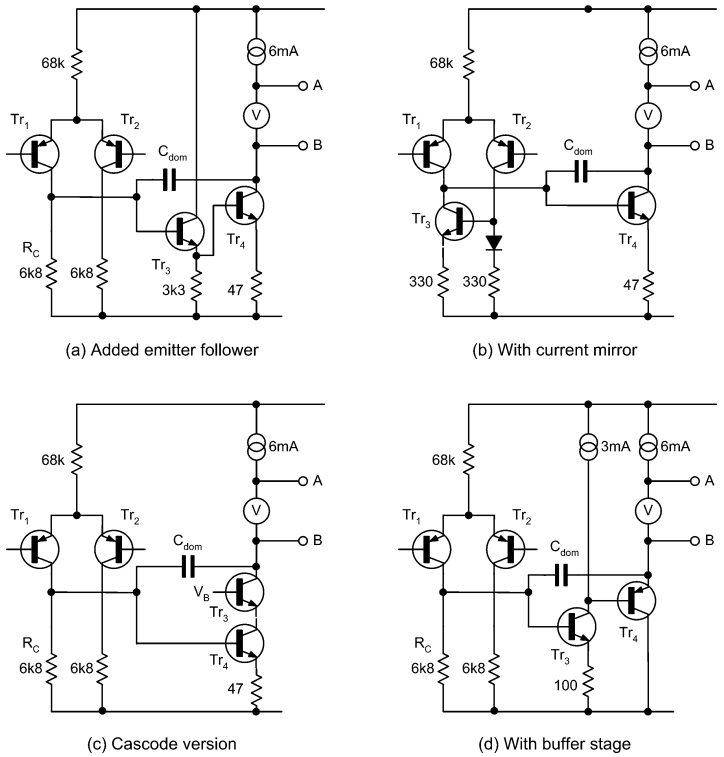



**Figure 4** *Enhancements to the basic amplifier configuration.*

The added emitter-follower and current-mirror enhancements (Figure 4a and b) work because the input pair act as a transconductance amplifier (voltage-difference in, current out) feeding a VAS that is basically a Miller integrator, thanks to the dominant-pole capacitance $C_{\mathrm{dom}}$.

Emitter-follower $\mathrm{Tr}_3$ increases open-loop gain by enhancing the current-gain of the VAS. The current mirror does the same thing by doubling input-pair transconductance; also slew-rate is greater and symmetrical ($30\,\mathrm{V}/\mu\mathrm{s}$ was obtained) as the input stage can now sink current as effectively as it sources it. Furthermore, input collector currents are kept balanced, valuable if the d.c. offset at the amplifier output is to be kept within acceptable bounds (say $\pm 50\,\mathrm{mV}$) without adjustment. The input collector currents, and hence the base currents drawn through input and feedback resistors, must be roughly equal.

Total input pair current is a critical parameter, since it affects input-stage transconductance and hence open-loop gain, and also defines the maximum slew rate, setting the maximum current that can flow in and out of the dominant-pole compensation capacitor $\mathrm{C}_{\mathrm{dom}}$. Though not obvious,

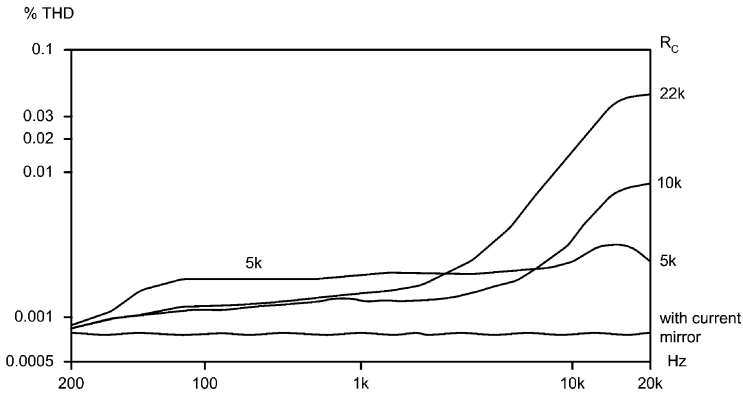



**Figure 5** *Effect of $R_c$ and current-mirror, input and VAS only.*

the input-pair current is an important influence on the HF stability of the amplifier. Power amplifier design has always been impeded by the fact that the crucial VAS, with its high-impedance collector, has to drive an output stage with markedly non-linear input impedance. The resulting interaction means that, when distortion occurs, it is not clear whether it arises in the output stage itself, or at the VAS due to non-linear loading.

Distortion not caused by the output can be studied by replacing the output stage with a very linear Class-A voltage-follower. Resistor $R_c$ has a strong effect on HF distortion (Figure 5) with a clear improvement when $R_c$ is replaced by a simple diode/transistor current mirror. This underlines the crucial nature of the interface between the input pair and the VAS.

Similarly, the output stage can be tested in isolation by driving it directly from a low-impedance oscillator[6] and this technique was used to study the two output arrangements described below.

Neither method, however, allows study of the VAS/output interaction directly. One strategy for side-stepping this problem is to use a unity-gain buffer between the VAS and the nonlinear load of the output stage (e.g. emitter-follower $Tr_4$, Figure 4d).

This approach has not been pursued with mosfets, but I have used it in a commercial design for a bipolar quasi-complementary amplifier. It proved effective at reducing distortion and is an approach which seems unfairly neglected.

## Reducing costs

In the search for a better amplifier, two routes were examined. The first is to reduce the cost of a mosfet power amp by using two N-channel devices in a form of quasi-complementary output stage.

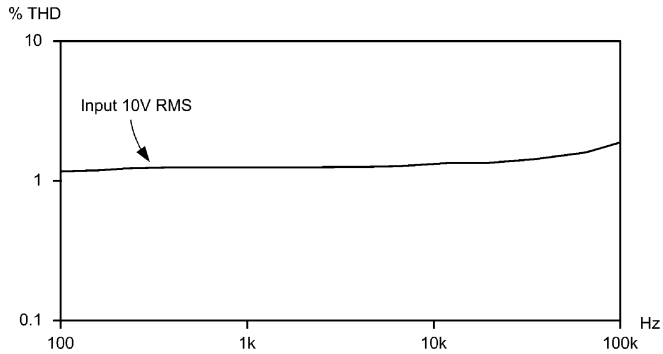



**Figure 6** *Quasi-complementary output stage only (open-loop) showing almost flat THD.* 8 Ω *resistive, measurement bandwidth* 22–7500 Hz.

The second is to increase linearity and improve quiescent stability by using bipolar drivers with local feedback around each output FET.

The quasi-complementary approach is directly analogous to that used in the early days of transistor amplifiers; a P-N-P transistor is combined with an N-channel FET to emulate a P-channel FET, and this works very well.

The output devices are now the same (promising bulk-buying economy) and can be chosen without reference to complements. That used was the IRF530, offering 100 V, 14 A, and 75 W in a TO-220 package, a pair costing £5.00 against £9.50 for the 2SK134/2SJ49 pair. Open-loop distortion of the output stage alone driven from a low impedance was shown (Figure 6) to be 1.1%.

Closed-loop performance (Figure 7), (yielded by the practical circuit shown in Figure 8) demonstrated that, predictably, output symmetry is not wonderful, and crossover effects on the residual were clear.

But this performance is acceptable from such a simple and economical circuit, and could almost certainly be improved at minor cost by adopting added emitter-follower and/or current mirror.

Tests showed that the value of $R_c$ in Figure 8 is non-critical, so there seems no difficulty in driving the bottom mosfet gate capacitance.

Quiescent setting is by transistor bias-generator; this is purely a regulator, and is *not* thermally coupled to the output devices. In an attempt to improve output symmetry, a Baxandall diode[10] was inserted at point D in the driver emitter. Sadly, THD was unchanged, despite the dependable improvement that this modification gives in bipolar quasi-complementary designs.

## Reducing distortion

If performance outweighs economy, a true complementary output pair is retained, with local feedback linearising each mosfet.[6] An important

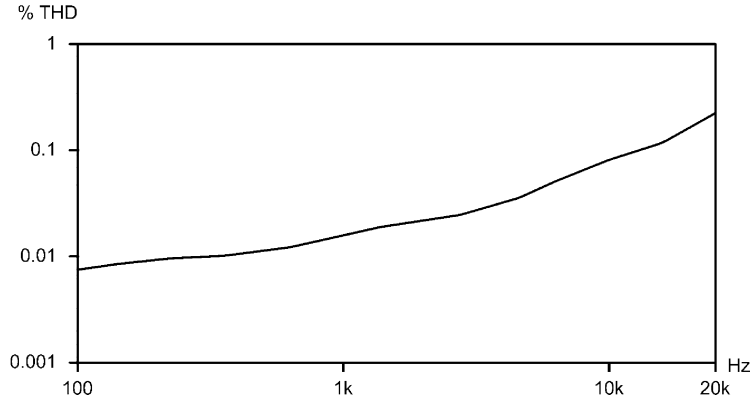



**Figure 7** *Complete quasi-complementary amplifier (closed-loop).* 10 V *r.m.s. into* 8 Ω. *(Bandwidth* 100–80 kHz.*)*

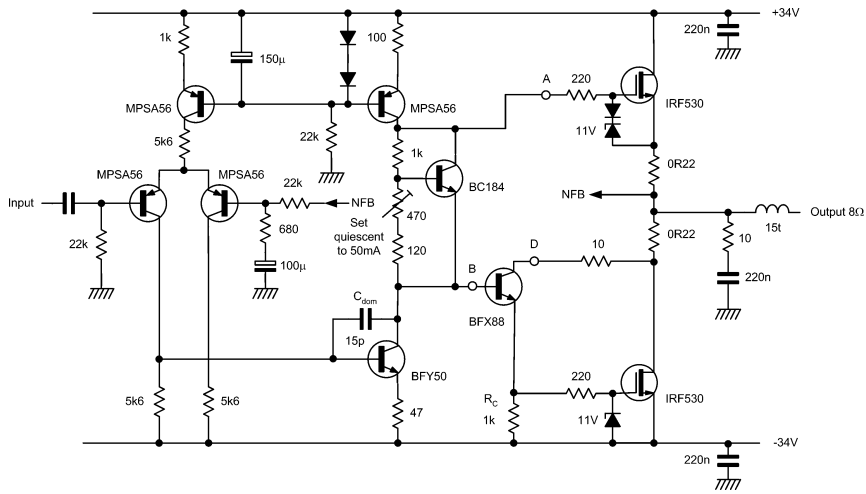



**Figure 8** *Practical circuit for quasi-complementary amplifier.*

variable here is the value of emitter-resistor $R_e$ in Figure 11. A high value increases output distortion, since it reduces the feedback factor within each hybrid bipolar-mosfet loop, while a low value makes quiescent-current setting unduly critical. As a compromise, 10–22 Ω works well.

Figure 9 shows the output stage alone in open loop giving 0.05%, and Figure 10 shows closed-loop performance as given by the practical design in Figure 11, plus the effect of varying $C_{dom}$.

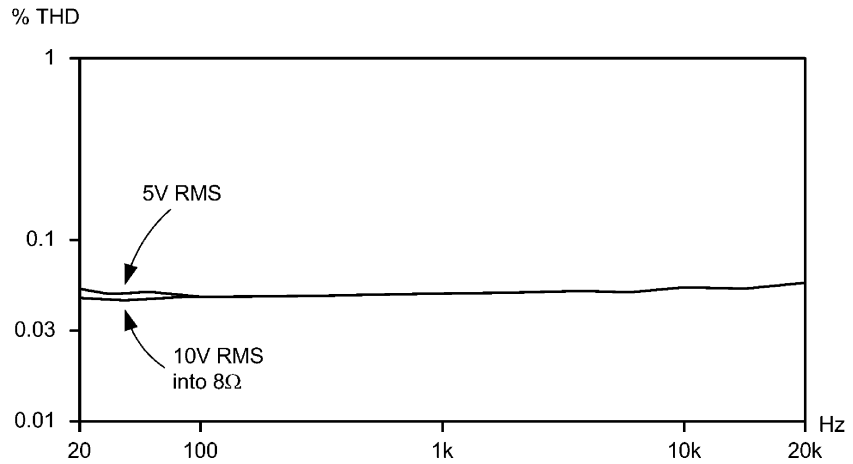



**Figure 9** *Hybrid output stage only, operated open-loop.*

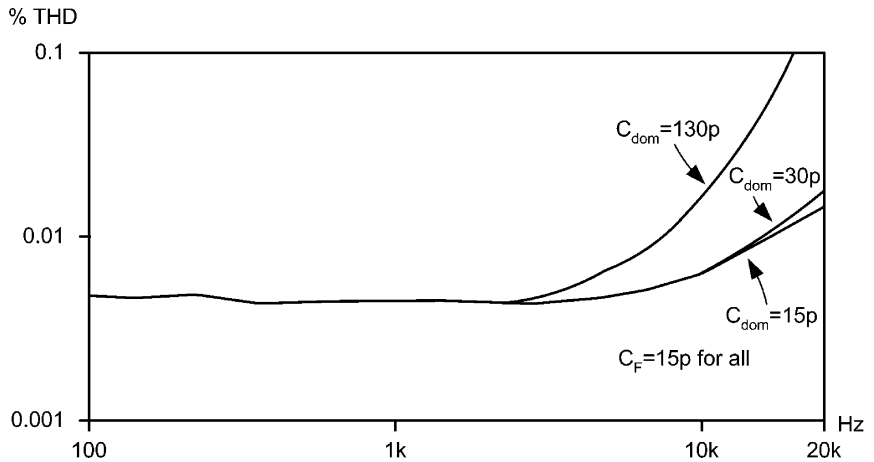



**Figure 10** *Hybrid amplifier (closed-loop) with varying values of $C_{dom}$. 16 V r.m.s. into 8 Ω.*

An emitter-follower is added to increase the feedback factor, and $C_f$ is now needed for HF stability with $C_{dom} = 15$ pF. The compensation is not necessarily optimised for all possible real-life loads.

Distortion across the band is now very low. While it is desirable to define the closed-loop bandwidth of any audio device, do not put a simple RC filter at the input, as this makes the bandwidth dependent on the output impedance of the upstream equipment.

The temptation to implement it by increasing $C_f$ should also be resisted, because if it is large enough to provide a suitable roll-off, there is a danger

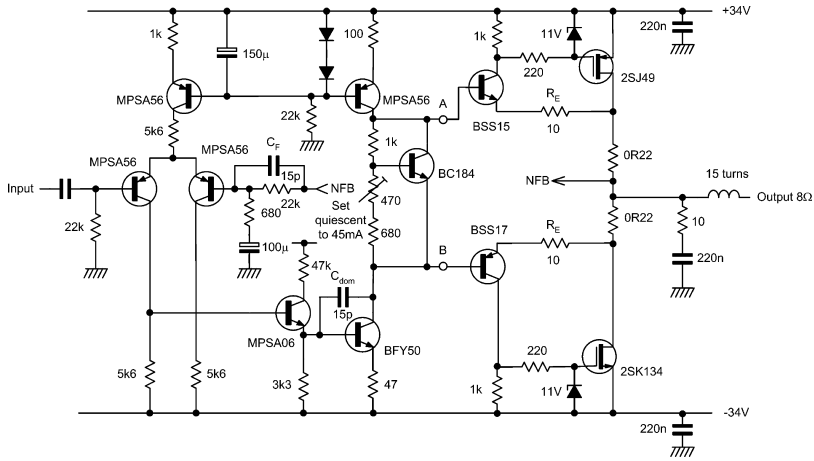



**Figure 11** *Practical circuit for high-quality hybrid bipolar-mosfet amplifier.*

that it will induce mysterious VHF instability in $Tr_2$ – a common effect. It is only revealed by distortion that vanishes at the touch of a cautious finger.

I have not tried to give the final word on this subject, though I hope it illuminates a few new directions. The use of current mirror topology in particular looks promising.

## References


1. Kessler, K. 'Valfet Amplifier', *HiFi News and Record Review*, June 1989, p 65.
2. Self, D. 'Electrolytics & Distortion', *EW + WW*, February 1989, pp 43–44.
3. Linsley Hood, J. 'Class AB Mosfet Power Amp', *EW + WW*, March 1989, p 264.
4. Self, D. 'Science vs Subjectivism', *EW & WW*, July 1988, p 695.
5. Brown, I. 'Feedback and FETs in Audio Power Amps', *EW + WW*, February 1989, pp 123–124.
6. Self, D. 'Mosfet Audio Output', Letter, *EW + WW*, May 1989, p 514.
7. Borbely, E. 'High Power Amplifier', *Wireless World*, March 1983, pp 69–75.
8. Tilbrook, D. '150W Mosfet Amplifier', *ETI*, June 1982, pp 48–55.
9. Cordell, R. 'Mosfet Power Amplifier with Error-Correction', *JAES*, January/February 1984, pp 6–17.
10. Baxandall, P. 'Symmetry in Class B', Letter, *Wireless World*, September 1969, p 416.

# 15 FETs versus BJTs: the linearity competition

*May 1995*

This short chapter was published when debate about the linearity of power FETs was raging, or at any rate smouldering, in the letters columns of Electronics World. Many contributors were content to point out that FETs must be more linear than bipolar transistors because everyone says so. This sort of argument has never had much appeal for me, and I carried out several investigations to see if there was any way in which FETs could be claimed to produce less distortion. One of the problems with this is making a meaningful comparison between two rather different kinds of active device. Here I try to level the playing field by making the transconductances the same; the BJT wins heavily on linearity when degenerated to have the same low transconductance as a FET. In a complete power amplifier, the situation is naturally rather more complex; BJT output devices need BJT drivers (I suppose you could use FET drivers, but I think it a most unpromising route to head down) which introduce more distortion than you might expect, while many FET power output stages can dispense with drivers altogether, so long as the VAS is capable of charging and discharging those rather large gate-capacitances. Nevertheless, I am confident that in a fair contest a BJT amplifier will always have lower distortion than its FET equivalent. In particular, it will have lower and less nasty crossover artifacts.

Not everyone felt that this contribution settled the matter for good – the SPICE models used to simulate the FETs were the focus of particular attention. Fortunately, since then a lot more

has been published on FET models, and it appears my conclusions were correct.

There has been much debate recently as to whether power FETs or bipolar junction transistors (BJTs) are superior in power amplifier output stages. Reference 1 is a good example. It has often been asserted that power FETs are more linear than BJTs, usually in tones that suggest that only the truly benighted are unaware of this.

In audio electronics it is a good rule of thumb that if an apparent fact is repeated times without number, but also without any supporting data, it needs to be looked at very carefully indeed. I therefore present my own view of the situation here, in the hope that the resulting heat may generate some light.

I suggest that it is now well-established that power FETs, when used in conventional Class-B output stages, are a good deal less linear than BJTs.[2] Gain deviations around the crossover region are far more severe for FETs than the relatively modest wobbles of correctly biased BJTs, and the shape of the FET gain-plot is inherently jagged, due to the way in which two square-law devices overlap.

The incremental gain range of a simple FET output stage is 0.84 to 0.79, range 0.05, and this is actually much greater than for the bipolar stages in Ref. 2; the emitter-follower stage gives 0.965 to 0.972 into 8 Ω, with a range of 0.007, and the complementary feedback pair gives 0.967 to 0.970 with a range of 0.003. The smaller ranges of gain-variation are reflected in the much lower THD figures when PSpice data is subjected to Fourier analysis.

However, the most important difference may be that the bipolar gain variations are gentle wobbles, while all FET plots seem to have abrupt changes. These are much harder to linearise with negative feedback that must decline with rising frequency. The basically exponential $I_c/V_{be}$ characteristics of two BJTs approach much more closely the ideal of conjugate mathematical functions – i.e. always adding up to 1. This is the root cause of the much lower crossover distortion.

Close-up examination of the way in which the two types of device begin conducting as their input voltages increase shows that FETs move abruptly into the square-law part of their characteristic, while the exponential behaviour of bipolar devices actually gives a much slower and smoother start to conduction (see Figures 4 and 5).

Similarly, recent work shows that less conventional approaches, such as the common-collector/common-emitter configuration of Bengt Olsson, also suffer from the non-conjugate nature of FETs. They also show sharp changes in gain. Gevel[3] shows that this holds for both versions of the stage proposed by Olsson, using both N and P-channel drivers. There are always sharp gain-changes.

## Class A stage

It occurred to me that the idea that FETs are more linear was based not on Class-B power-amplifier applications, but on the behaviour of a single device in Class-A. You might argue that the roughly square-law nature of a fet's $I_d/V_{gs}$ law is intuitively more 'linear' than the exponential $I_c/V_{be}$ law of a BJT, but it is difficult to know quite how to define 'linear' in this context. Certainly a square-law device will generate predominantly low-order harmonics, but this says nothing about the relative amounts produced.

In truth the BJT/FET contest is a comparison between apples and aardvarks, the main problem being that the raw transconductance ($g_m$) of a BJT is far higher than for any power FET. Figure 1 illustrates the conceptual test circuit; both a TO3 BJT *MJ802* and an *IRF240* power FET have an increasing d.c. voltage, $V_{in}$, applied to their base/gate, and the resulting collector and drain currents from PSpice simulation are plotted in Figure 2.

Voltage $V_{offset}$ is used to increase the voltage applied to FET $M_1$ by 3.0 V because nothing much happens below a $V_{gs}$ of 4 V, and it is helpful to have the curves on roughly the same axis. Curve A, for the bjt, goes almost vertically skywards, as a result of its far higher $g_m$. To make the comparison meaningful, a small amount of local negative feedback is added to $Q_1$ by $R_e$. As this emitter degeneration is increased from 0.01 to 0.1 Ω, the $I_c$ curves become closer in slope to the $I_d$ curve.

Because of the curved nature of the FET $I_d$ plot, it is not possible to pick an $R_e$ value that allows very close $g_m$ equivalence; a value of 0.1 Ω was chosen for $R_e$, this being a reasonable approximation; see Curve B. However, the important point is that I think no-one could argue that the FET $I_d$ characteristic is more linear than Curve B.

This is made clearer by Figure 3, which directly plots transconductance against input voltage. There is no question that FET transconductance

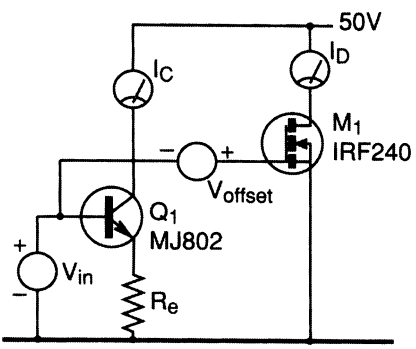



**Figure 1** *Linearity test circuit. Voltage* $V_{offset}$ *adds* 3 V *to the d.c. level applied to the FET gate, purely to keep the current curves helpfully adjacent on a graph.*

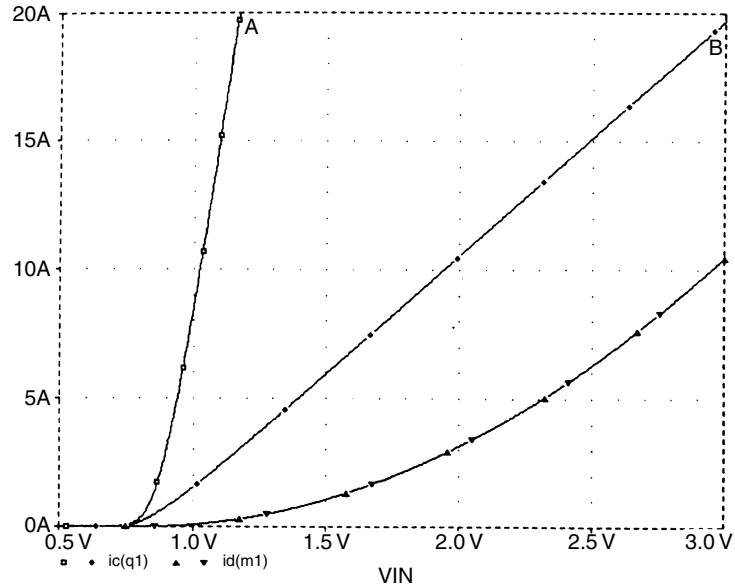



**Figure 2** *Graph of* $I_c$ *and* $I_d$ *for the BJT and the FET. Curve A shows* Ic *for the BJT alone, while Curve B is the result for* $R_e = 100\,m\,\Omega$. *The curved line is the* $I_d$ *result for a power FET without any degeneration.*

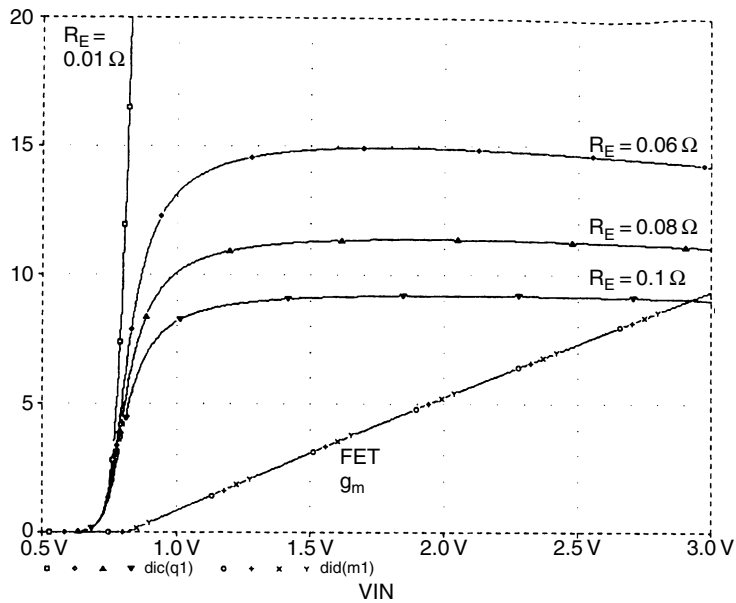



**Figure 3** *Graph of transconductance versus input voltage for BJT and FET. The near-horizontal lines are BJT* $g_m$ *for various* $R_E$ *values.*

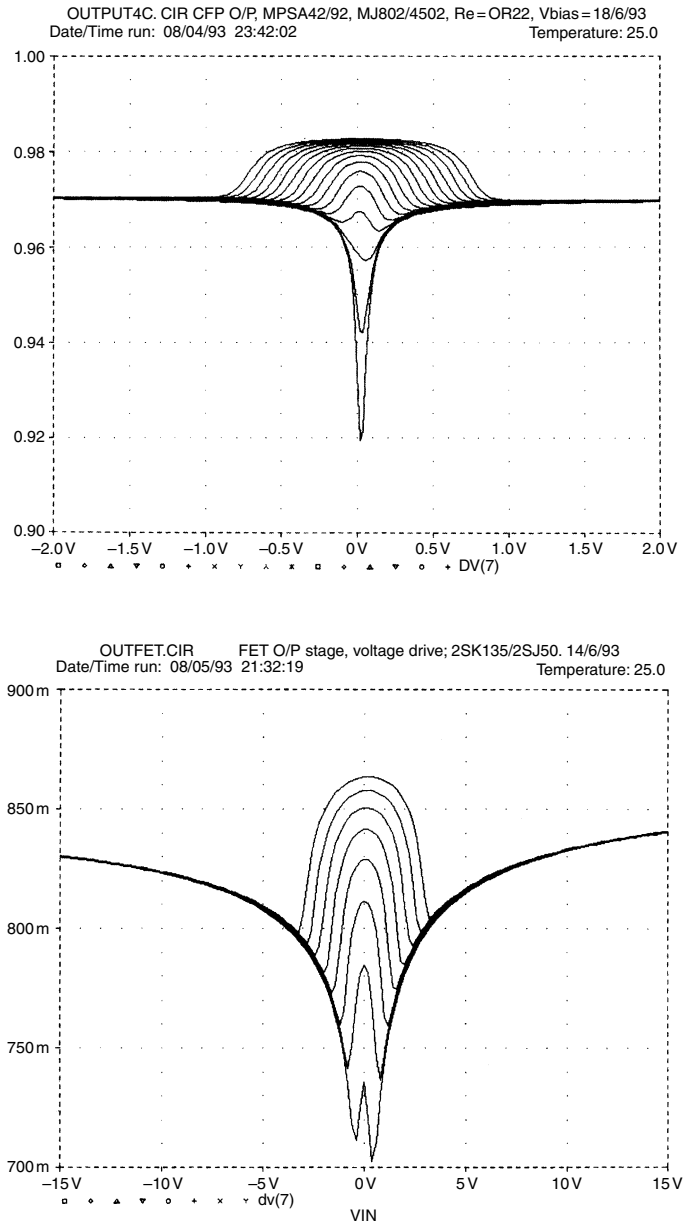



**Figure 4 and 5** *Top are curves for a bipolar complementary feedback pair, crossover region* ±2 V, $V_{bias}$ *as a parameter. Fourth curve up provides good optimal setting – compare with curves below, for a FET source follower crossover region with* ±15 V *range.*

increases in a beautifully linear manner-but this 'linearity' is what results in a square-law $I_d$ increase. The near-constant $g_m$ lines for the BJT are a much more promising basis for the design of a linear amplifier.

To forestall any objections that this comparison is nonsense because a BJT is a current-operated device, I add here a small reminder that this is untrue. The BJT is a voltage operated device, and the base current that flows is merely an inconvenient side-effect of the collector current induced by said base voltage. This is why beta varies more than most BJT parameters; the base current is an unavoidable error rather than the basis of transistor operation.

The PSpice simulation shown was checked against manufacturers' curves for the devices, and the agreement was very good – almost unnervingly so. It therefore seems reasonable to rely on simulator output for these kind of studies; it is certainly infinitely quicker than doing the real measurements. In addition, the comprehensive power-FET component libraries that are part of PSpice allow the testing to be generalised over a huge number of component types without you needing to buy them.

To conclude, I think it is probably irrelevant to simply compare a naked BJT with a naked FET. Perhaps the vital point is that a bipolar device has much more raw transconductance gain to begin with, and this can be handily converted into better linearity by local feedback, i.e. adding a little emitter degeneration.

If the transconductance is thus brought down roughly to FET levels, the bipolar has far superior large-signal linearity. I must admit to a sneaking feeling that if practical power BJTs had come along after FETs, they would have been seized upon with glee as a major step forward in power amplification.

## References

1. Hawtin, V. Letters, *EW+WW* December 1994, p 1037.
2. Self, D. 'Distortion in power amplifiers', Part 4, *EW + WW*, November 1993, pp 932–934.
3. Gevel, M. Private Communication, January 1995.

# 16 Distortion in power amplifiers, Part I: the sources of distortion

*August 1993*

For many years I felt that the output stages of power amplifiers presented very great possibilities for creative design, and I actually got round to exploring some of them. The hybrid bipolar-FET output stage, also collected in this volume, was one of them. One of the first difficulties I met was the problem of determining how much of the total distortion was due to the output stage, and how much was being produced in the small-signal sections. This latter contribution turned out to be larger and more important than I expected. Very little reliable information appeared to exist on amplifier distortion, and I found myself embarked on a major effort to track down all the sources of distortion in the typical solid-state power amplifier. The initial investigations were not very illuminating, until I realised that changing a component value in the typical power amplifier circuit very often alters two or more distortion mechanisms simultaneously, making the results hard to interpret. I was sitting in my armchair late one Spring night when the full force of this struck home, and thereafter I devised ways to simulate or measure the distortion mechanisms in isolation. This approach is well-illustrated in Parts 2 and 3 of the series, dealing with the input pair and the Voltage-Amplifier Stage, respectively.

Things then began to fall into place, and one day I put together all the various minor and, apparently insignificant, improvements I had made to the utterly conventional amplifier circuit I had started with. The distortion was exhilaratingly low, stability was good, and I soon felt that I could write a pretty comprehensive guide to distortion and power amplifier design. Electronics

> World were good enough to give me all the space I asked for, and I believe I succeeded; here, in eight chapters, is the result.

It seems surprising that in a world which can build the Space Shuttle and detect the echoes of the birth of the universe, we still have to tolerate distortion in power amplifiers. Leafing through recent reviews and specifications shows claims for full-power total harmonic distortion ranging more than three orders of magnitude between individual designs, a wider range than any other parameter.

Admittedly the higher end of this range is represented by subjectivist equipment that displays dire linearity, presumably with the intention of implying that other nameless audio properties have been given priority over the mundane business of getting the signal from input to output without bending it.

Given the juggernaut rate of progress in most branches of electronics this seems to me anomalous, and especially notable in view of the many advanced analogue techniques used in op-amp design; after all power amps are only op-amps with boots on. One conclusion seems inescapable: a lot of power amplifiers generate much more distortion than they need to.

This series attempts to show exactly why amplifiers distort, and how to stop them doing it, culminating in a practical design for an ultra-linear amplifier. It should perhaps be said at the outset that none of this depends on excessively high levels of negative feedback. Many of the techniques described here are also entirely applicable to discrete op-amps, headphone drivers, and similar circuit blocks. Since we are almost in the twenty-first century I have ignored valve amplifiers.

Since mis-statements and confusions are endemic to audio, I have based these articles almost entirely on my own experimental work backed up with spice circuit simulation; much of the material relates specifically to bipolar transistor output stages, though a good deal is also relevant to mosfet amplifiers. Some of the statements made may seem controversial, but I believe they are all correct. If you think not, please tell me, but only if you have some real evidence to offer.

The fundamental reason why amplifier distortion persists is, of course, because it is a difficult technical problem to solve. A Science proverbially becomes an Art when there are more than seven variables, and since it will emerge that there are seven major distortion mechanisms to the average amplifier, we would seem to be nicely balanced on the boundary of the two cultures. Given so many significant sources of unwanted harmonics, overlaid and sometimes partially cancelling, sorting them out is a non-trivial task.

*Make your amplifier as linear as possible before applying NFB* has long been a cliche, (one that conveniently ignores the difficulty of running a high

gain amp without any feedback) but virtually no dependable advice on how to perform this desirable linearisation has been published. The two factors are the basic linearity of the forward path, and the amount of negative feedback applied to further straighten it out. The latter cannot be increased beyond certain limits or high-frequency stability is put in peril, whereas there seems no reason why open-loop linearity could not, in principle, be improved without limit, leading us to the Holy Grail of the distortionless amplifier. This series therefore takes as its prime aim the understanding and improvement of open-loop linearity. As it proceeds we will accrete circuit blocks to culminate in two practical amplifier designs that exploit the techniques presented here.

## How an amplifier (really) works

Figure 1 shows the usual right trusty and well-beloved power amp circuit drawn as standard is possible. Much has been written about this configuration, though its subtlety and quiet effectiveness are usually overlooked, and the explanation below therefore touches on several aspects that seem to be almost unknown. It has the merit of being docile enough to be made into a workable amplifier by someone who has only the sketchiest of notions as to how it works.

The input differential pair implements one of the few forms of distortion cancellation that can be relied upon to keep working in all weathers. This is because the transconductance of the input pair is determined by the physics of transistor action rather than matching of variable parameters such as beta; the logarithmic relation between $I_c$ and $V_{be}$ is proverbially accurate over some eight or nine decades of current variation.

The voltage signal at the voltage amplifier stage (hereafter VAS) transistor base is typically a couple of millivolts, looking rather like a distorted triangle wave. Fortunately the voltage here is of little more than academic interest, as the circuit topology essentially consists of a transconductance amp (voltage-difference input to current output) driving into a transresistance (current-to-voltage converter) stage. In the first case the exponential $V_{be}/I_c$ law is straightened out by the differential-pair action, and in the second the global (overall) feedback factor at LF is sufficient to linearise the VAS, while at HF shunt negative feedback (hereafter NFB) through $C_{dom}$ conveniently takes over VAS-linearisation while the overall feedback factor is falling. The behaviour of Miller dominant-pole compensation in this stage is exceedingly elegant, and not at all just a case of finding the most vulnerable transistor and slugging it. As frequency rises and $C_{dom}$ begins to take effect, negative feedback is no longer applied globally around the whole amplifier, which would include the higher poles, but instead is seamlessly transferred to a purely local role in linearising

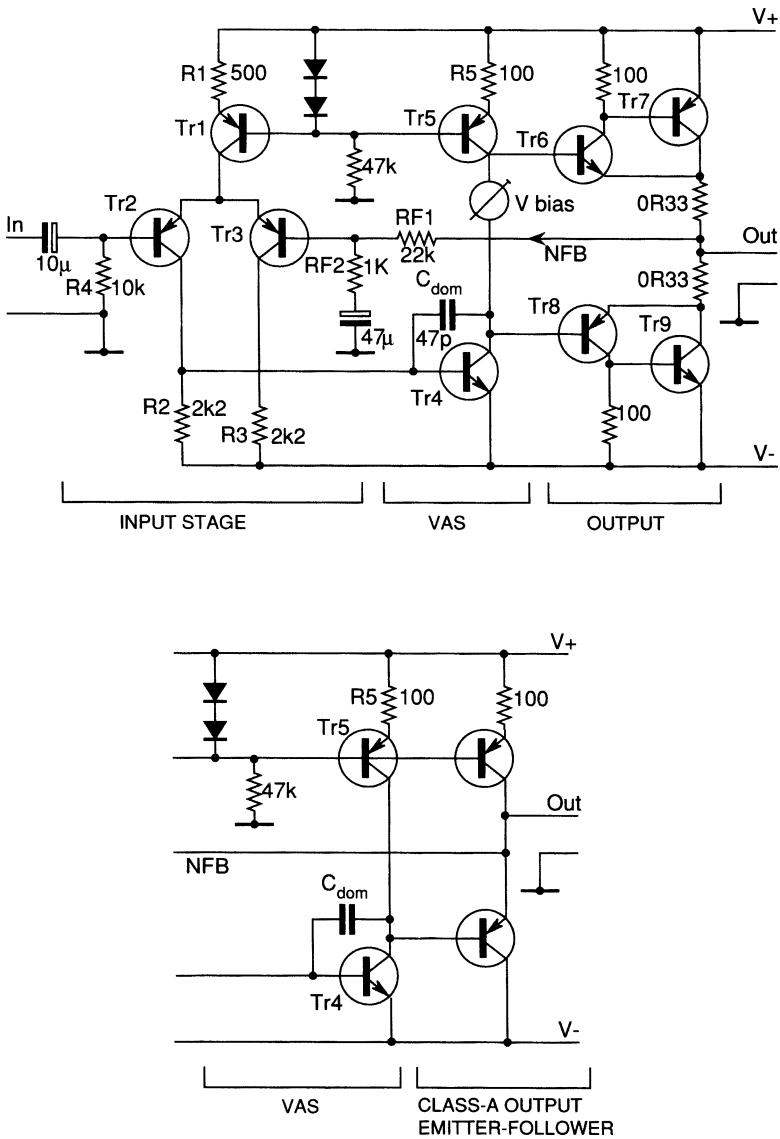



**Figure 1(a)** *Conventional class-B power amp circuit. The apparent simplicity of circuitry conceals a series of sophisticated operating mechanisms. The lower drawing (***Figure 1(b)***) shows an adaptation of the output stage for small signal modelling.*

the VAS. Since this stage effectively contains a single gain transistor, any amount of NFB can be applied to it without stability problems.

The amplifier operates in two regions; the LF, where open-loop gain is substantially constant, and HF, above the dominant-pole breakpoint,

where the gain is decreasing steadily at 6 dB/octave. Assuming the output stage is unity-gain, three simple relationships define the gain in these two regions:

$$\text{LF gain} = g_m \times \text{beta} \times R_c \tag{1}$$

At least one of the factors that set this (beta) is not well-controlled and so the LF gain of the amplifier is to a certain extent a matter of potluck; fortunately this doesn't matter as long as it is high enough to give a suitable level of NFB to eliminate LF distortion. The use of the word 'eliminate' is deliberate, as will be seen later. Usually the LF gain, or HF local feedback-factor, is made high by increasing the effective value of the VAS collector impedance $R_c$, either by the use of a currentsource collector-load, or by some form of bootstrapping.

The other important relations are:

$$\text{HF gain} = g_m/(\omega \times C_{dom}) \tag{2}$$

$$\text{Dominant pole freq } P1 = \frac{1}{(\omega \cdot C_{dom} \cdot \beta \cdot R_c)} \tag{3}$$

where $\omega = 2\pi f$.

In the HF region, things are distinctly more difficult as regards distortion, for while the VAS is locally linearised, the global feedback-factor available to linearise the input and output stages is falling steadily at 6 dB/octave. For the time being we will assume that it is possible to define an HF gain (say $N$dB at 20 kHz) which will assure stability with practical loads and component variations. Note that the HF gain, and therefore both HF distortion and stability margin, are set by the simple combination of the input stage transconductance and one capacitor, and most components have no effect on it at all.

It is often said that the use of a high VAS collector impedance provides a current drive to the output devices, often with the implication that this somehow allows the stage to skip quickly and lightly over the dreaded crossover region. This is a misconception – the collector impedance falls to a few kΩ at HF, due to increasing local feedback through $C_{dom}$. In any case it is very doubtful if true current drive would be a good thing since calculation shows that a low-impedance voltage drive minimises distortion due to beta-unmatched output halves,[1] and it certainly eliminates distortion mechanism four described later.

## The seven distortions

In the typical amplifier THD is often thought to be simply due to the Class-B nature of the output stage, which is linearised less effectively as the

feedback factor falls with increasing frequency. However the true situation is much more complex as the small-signal stages can generate significant distortion in their own right in at least two different ways. This can easily exceed the output stage distortion at high frequencies. It seems inept to allow this to occur given the freedom of design possible in the small-signal section.

Include all the ills that a class-B stage is prone to and then there are seven major distortion mechanisms.

Distortion in power amplifiers arises from:

1 *Non-linearity in the input stage.* If this is a carefully-balanced differential pair then distortion is typically only measurable at HF, rises at 18 dB/octave, and is almost pure third harmonic.

   If the input pair is unbalanced (which from published circuitry it usually is) then the HF distortion emerges from the noise floor earlier. As frequency increases, it rises at 12 dB/octave as it is mostly second harmonic.

2 *Non-linearity in the voltage amplifier stage* surprisingly does not always figure in the total distortion. If it does, it remains constant until the dominant-pole frequency P1 is reached, and then rises at 6 dB/octave. With the configurations discussed here, it is always second harmonic.

   Usually the level is very low due to linearising negative feedback through the dominant-pole capacitor. Hence if you crank up the *local* VAS open-loop gain, for example by cascoding or putting more current-gain into the local VAS/$C_{dom}$ loop, and attend to mechanism four below, you can usually ignore VAS distortion.

3 *Non-linearity in the output stage*, which is naturally the obvious source. This, in a Class-B amplifier, will be a complex mix of largesignal distortion and crossover effects, the latter generating a spray of high-order harmonics, and in general rising at 6 dB/octave as the amount of negative feedback decreases. Large-signal THD worsens with 4 Ω loads and worsens again at 2 Ω. The picture is complicated by dilatory switch-off in the relatively slow output devices, ominously signalled by supply current increasing in the top audio octaves.

4 *Loading of the VAS by the non-linear input impedance of the output stage.* When all other distortion sources have been attended to, this is the limiting distortion factor at LF (say below 2 kHz). It is simply cured by buffering the VAS from the output stage. Magnitude is essentially constant with frequency, though overall effect in a complete amplifier becomes less as frequency rises and feedback through $C_{dom}$ starts to linearise the VAS.

5 *Non-linearity caused by large rail-decoupling capacitors feeding the distorted signals on the supply lines into the signal ground.* This seems to be the reason many amplifiers have rising THD at low frequencies. Examining one

commercial amplifier kit, I found that rerouting the decoupler ground-return reduced THD at 20 Hz by a factor of three.

6 *Non-linearity caused by induction of Class-B supply currents into the output, ground, or negative-feedback lines.* This was highlighted by Cherry[3] but seems to remain largely unknown; it is an insidious distortion that is hard to remove, though when you know what to look for on the THD residual, it is fairly easy to identify. I suspect that a large number of commercial amplifiers suffer from this to some extent.

7 *Non-linearity resulting from taking the NFB feed from slightly the wrong place near where the power-transistor Class-B currents sum to form the output.* This may well be another common defect.

Having set down what Mao might have called The Seven Great Distortions – Figure 2 shows the location of these mechanisms diagrammatically – we may pause to put to flight a few Paper Tigers. The first is common-mode distortion in the input stage, a spectre that tends to haunt the correspondence columns. Since it is fairly easy to make an amplifier with less than $<0.00065\%$ THD (1 kHz) without paying any special attention to this, it cannot be too serious a problem. A more severe test is to apply the full output voltage as a common-mode signal, by running the amplifier as a unity-gain voltage-follower. If this is done using a model (see below for explanation) small-signal version of Figure 1, with suitable

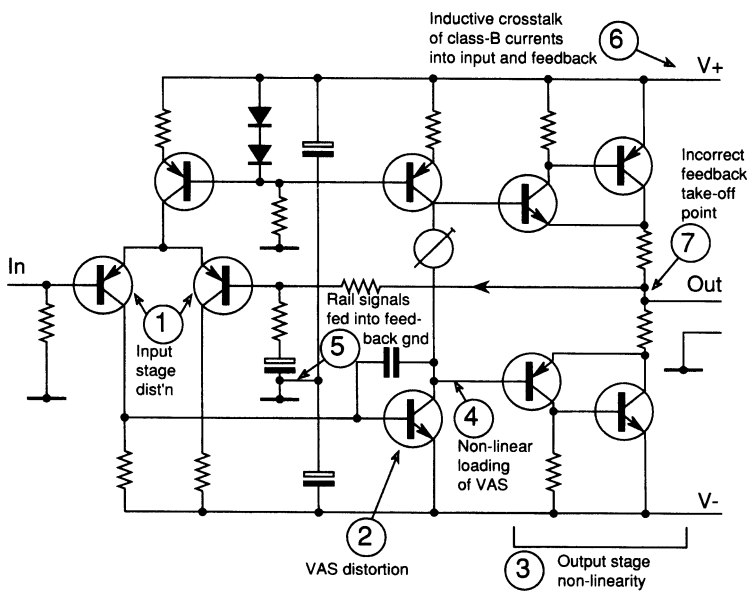



**Figure 2** *A topology of distortion: the location of the seven distortion mechanisms.*

attention to compensation, then it yields less than 0.001% at 8V r.m.s. across the audio band. It therefore appears that the only real precaution required against common-mode distortion is to use a tail current-source for the input pair.

The second distortion conspicuous by its absence in the list is the injection of distorted supply-rail signals directly into the amplifier circuitry. Although this putative mechanism has received a lot of attention,[4] dealing with Distortion five above by proper grounding seems to be all that is required ... Once again, if triple-zero THD can be attained using simple unregulated supplies and without specifically addressing power supply rejection ratio, (which it reliably can be) then much of the work done on regulated supplies may be of doubtful utility. However, PSRR does need some attention if the hum/noise performance is to be of the first order.

A third mechanism of doubtful validity is thermal distortion, allegedly induced by parameter changes in semiconductor devices whose instantaneous power dissipation varies over a cycle. This would presumably manifest itself as a distortion increase at very low frequencies, but it simply does not seem to happen. The major effects would be expected in Class-B output stages where dissipation can vary wildly over a cycle. However drivers and output devices have relatively large junctions with high thermal inertia. Low frequencies are of course also where the NFB factor is at its maximum.

## The advantages of being conventional

The input pair not only provides the simplest way of making a d.c. coupled amp with a dependably small output offset voltage, but can also (given half a chance) completely cancel the second-harmonic distortion which would be generated by a single-transistor input stage. One vital condition must be met; the pair must be accurately balanced by choosing the associated components so that the two collector currents are equal. (The 'typical' component values shown in Figure 1 *do not* bring about this most desirable state of affairs.)

The input devices work at a constant and near-equal $V_{ce}$, giving good thermal balance.

The input pair has virtually no voltage gain so no low-frequency pole can be generated by Miller effect in the $Tr_2$ collector-base capacitance. All the voltage gain is provided by the VAS stage, which makes for easy compensation. Feedback through $C_{dom}$ lowers VAS input and output impedances, minimising the effect of input-stage and output stage capacitance. This is often known as pole-splitting;[2] the pole of the VAS

is moved downwards in frequency to become the dominant pole, while the input-stage pole is pushed up in frequency.

The VAS Miller compensation capacitance smoothly transfers NFB from a global loop which may be unstable, to the VAS local loop that cannot be. It is quite wrong to state that *all* the benefits of feedback are lost as the frequency increases above the dominant pole, as the VAS is still being linearised. This position of $C_{dom}$ also swamps the rather variable $C_{cb}$ of the VAS transistor.

To return to our list of the unmagnificent seven, note that only Distortion three is *directly* due to O/P stage non-linearity, though numbers 4–7 all result from the Class-B nature of the typical output stage.

## The performance

The THD curve for the standard amplifier is shown in Figure 3. As usual the distortion increases with frequency and, as we shall see later, would give grounds for suspicion if it did not. The flat part of the curve below 500 Hz represents non-frequency-sensitive distortion rather than the noise floor, which for this case is at about the 0.0005% level. Above 500 Hz the distortion rises at an increasing rate, rather than a constant number of dB/octave, due to the combination of Distortions 1,2,3 and 4. (In this case Distortions 5,6 and 7 have been carefully eliminated to keep things

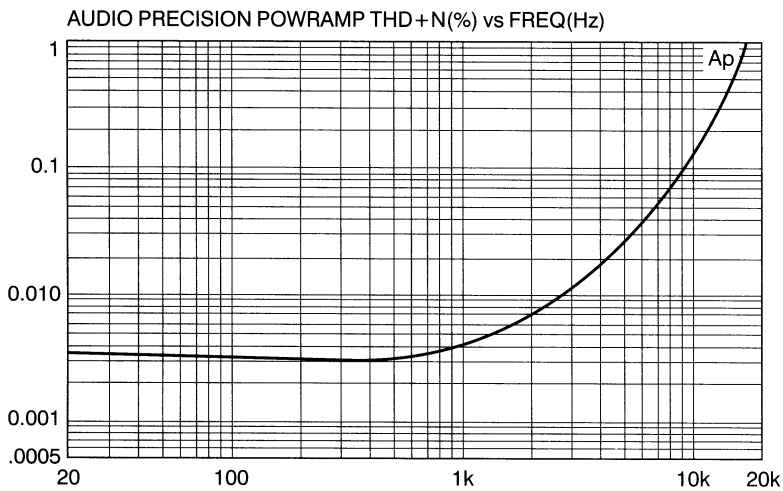



**Figure 3** *The distortion performance of the class-B amplifier shown in Figure 1(a).*

simple. This is why the distortion performance looks good already, and the significance of this should not be overlooked.) It is often written that having distortion constant across the audio band is a good thing. This is a most unhappy conclusion as the only practical way to achieve this with a Class-B amplifier is to *increase* the distortion at LF, for example by allowing the VAS to distort significantly.

It should now be clear why it can be hard to wring linearity out of a snake-pit of contending distortions. A circuit-value change is likely to alter at least two of the distortion mechanisms, and probably change the open-loop gain as well. In the coming articles I shall demonstrate how each of these mechanisms can be measured and manipulated separately.

## Determining open-loop linearity

Improving something demands its measurement, and so it is essential to examine the open-loop linearity of typical power-amp circuits. This cannot in general be done directly, so it is necessary to measure the NFB factor and calculate open-loop distortion from the usual closed-loop data. It is assumed that the closed-loop gain is fixed by operational requirements.

Finding the feedback-factor is at first sight difficult, as it means determining the open-loop gain. The standard methods for measuring op-amp open-loop gain involve breaking feedback-loops and manipulating closed-loop gains, procedures that are likely to send the average power-amplifier into fits. However, the need to measure this parameter is inescapable, as a typical circuit modification – e.g. changing the value of $R_2$ – will change the open-loop gain as well as the linearity, and to prevent total confusion it is necessary to keep a very clear idea of whether the observed change is due to an improvement in open-loop linearity or merely because the open-loop gain has risen. It is wise to keep a running check on the feedback-factor as work proceeds, and so the direct method of open-loop gain measurement shown in Figure 4 was evolved.

## Direct open-loop gain measurement

Since the amplifier shown in Figure 1 is a differential amplifier, its open-loop gain is simply the output divided by the voltage difference between the inputs. If the output voltage is kept effectively constant by providing a swept-frequency constant voltage at the +ve input, then a plot of open-loop gain versus frequency is obtained by measuring the error-voltage between the inputs, and referring this to the output level. This yields an upside-down plot that rises at HF rather than falling, as the differential amplifier

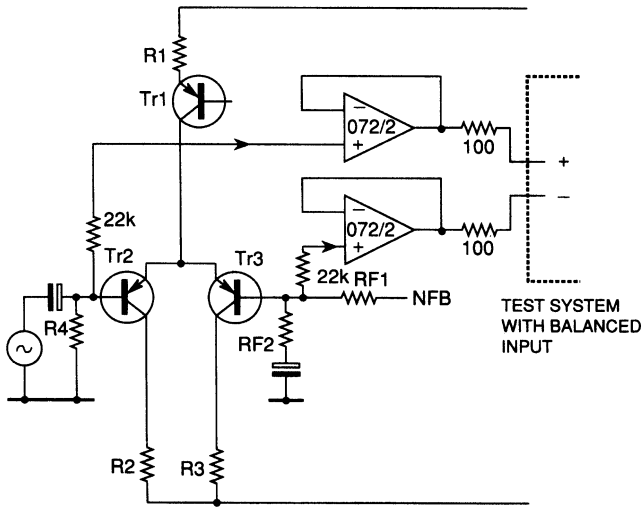



**Figure 4** *Test circuit for measuring open-loop gain directly. The measurement accuracy depends on the test gear CMRR.*

requires more input for the same output as frequency increases, but the method is so quick and convenient that this can be lived with. Gain is plotted in dB with respect to the chosen output level (+16 dBu in this case) and the actual gain at any frequency can be read off simply by dropping the minus sign. Figure 5 shows the plot for the amplifier in Figure 1.

The HF-region gain slope is always 6 dB/octave unless you are using something special in the way of compensation and, by the Nyquist rules, must continue at this slope until it intersects the horizontal line representing the feedback factor provided that the amplifier is stable. In other words, the slope is not being accelerated by other poles until the loop gain has fallen to unity, and this provides a simple way of putting a lower bound on the next pole P2; the important P2 frequency (which is usually somewhat mysterious) must be above the intersection frequency if the amplifier is seen to be stable.

Given test gear with a sufficiently high common-mode-rejection-ratio balanced input, the method of Figure 4 is simple; just buffer the differential inputs from the cable capacitance with *TL072* buffers, placing negligible loading on the circuit if normal component values are used. Be particularly wary of adding stray capacitance to ground to the −ve input, as this directly imperils amplifier stability by adding an extra feedback pole. Short wires from power amplifier to buffer IC can usually be unscreened as they are driven from low impedances.

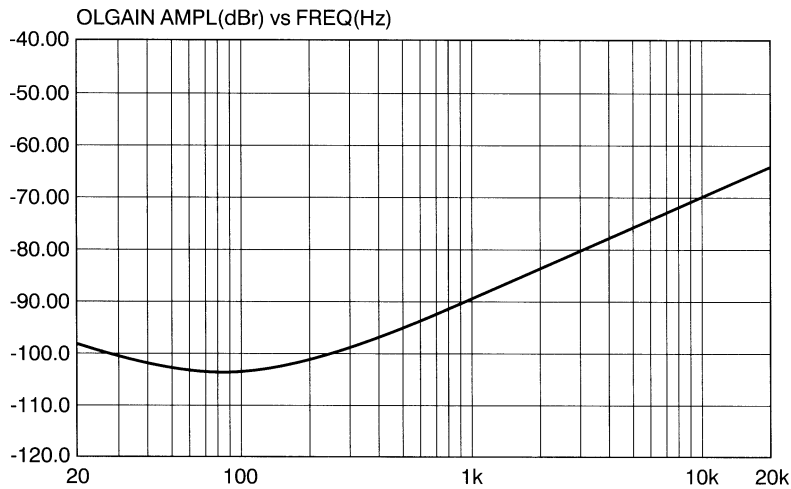



**Figure 5** *Open-loop gain versus frequency plot for Figure 1. Note that the curve rises as gain falls, because the amplifier error is the actual quantity measured.*

The test gear input CMRR defines the maximum open-loop gain measurable; I used an Audio Precision System-1 without any special alignment of CMRR. A calibration plot can be produced by feeding the two buffer inputs from the same signal; this will probably be found to rise at 6 dB/octave, being set by the inevitable input asymmetry. This must be low enough for amplifier error signals to be above it by at least 10 dB for reasonable accuracy. The calibration plot will flatten out at low frequencies, and may even show an LF rise due to imbalance of the test gear input-blocking capacitors; this can make determination of the lowest pole P1 difficult, but this is not usually a vital parameter in itself.

## Model amplifiers

The first two distortions on the list can dominate amplifier performance and need to be studied without the complications introduced by a Class-B output stage. This can be done by reducing the circuit to a model amplifier that consists of the small-signal stages alone, with a very linear Class A emitter-follower attached to the output to allow driving the feedback network. Here 'small-signal' refers to current rather than voltage, as the model amplifier should be capable of giving a full power-amp voltage swing, given sufficiently high rail voltages. From Figure 2 it is clear that this will allow study of Distortions 1 and 2 in isolation, and using this approach it will

prove relatively easy to design a small-signal amplifier with negligible distortion across the audio band. This is the only sure foundation on which to build a good power amplifier.

A typical plot combining Distortions 1 and 2 from a model amp is shown in Figure 6, where it can be seen that the distortion rises with an accelerating slope, as the initial rise at 6 dB/octave from the VAS is contributed to and then dominated by the 12 dB/octave rise in distortion from an unbalanced input stage.

The model can be powered from a regulated current-limited PSU to cut down the number of variables, and a standard output level chosen for comparison of different amplifier configurations. The rails and output level used for the results in these articles was ±15 V and +16 dBu. The rail voltages can be made comfortably lower than the average amplifier HT rail, so that radical bits of circuitry can be tried out without the creation of a silicon cemetery around your feet. It must be remembered that some phenomena such as input-pair distortion depend on absolute output level, rather than the proportion of the rail voltage used in the output swing, and will be worse by a mathematically predictable amount when the real voltage swings are used. The use of such model amplifiers requires some caution, and cannot be applied to bipolar output stages whose behaviour is heavily influenced by the sloth and low current gain of the power devices. As another general rule, if it is not possible to lash on a real output stage quickly and get a stable and workable power amplifier; the model may be dangerously unrealistic.

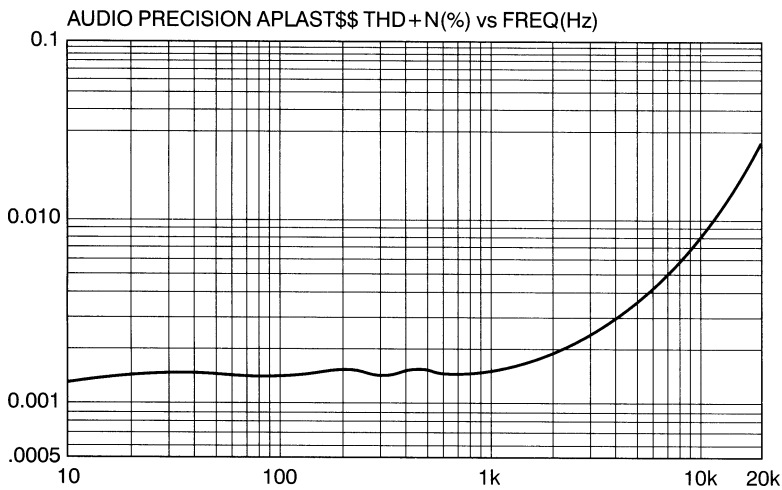



**Figure 6** *The distortion from a model amplifier, produced by the input pair and the voltage amplifier stage. Note increasing slope as input pair distortion begins to add to VAS distortion.*

## Glossary

Several abbreviations will be used throughout this series to keep its length under control.

**l.f.** Relating to amplifier action below the dominant pole, where the open-loop gain is assumed to be essentially flat with frequency.
**h.f.** Amplifier behaviour above the dominant pole frequency, where the open-loop gain is usually falling at 6 dB/octave.
**i/p** Input.
**p1** The first open-loop response pole, and its frequency in Hz.
**nfb** Negative feedback.

## References

1. Oliver *Distortion in Complementary-Pair Class-B Amplifiers*, Hewlett-Packard Journal, February 1971, p 11.
2. Feucht *Handbook of Analog Circuit Design*, Academic Press 1990, p 256 (Pole-splitting).
3. Cherry 'A new distortion mechanism in class-B amplifiers', *JAES*, May 1981, p 327.
4. Duncan 'PSU regulation boosts audio performance', *EW + WW*, October 1992, p 818.

# 17 Distortion in power amplifiers, Part II: the input stage

*September 1993*

The input stage of an amplifier performs the critical duty of subtracting the feedback signal from the input, to generate the error signal that drives the output. It is almost invariably a differential transconductance stage; a voltage-difference input results in a current output that is essentially insensitive to the voltage at the output port. Its design is also frequently neglected, as it is assumed that the signals involved must be small, and that its linearity can therefore be taken lightly compared with that of the voltage amplifier stage (VAS) or the output stage. This is quite wrong, for a misconceived or even mildly wayward input stage can easily dominate HF distortion performance.

The input transconductance is one of the two parameters setting HF open-loop (o/l) gain, and thus has a powerful influence on stability and transient behaviour as well as distortion. Ideally the designer should set out with some notion of how much o/l gain at 20 kHz will be safe when driving worst-case reactive loads – a precise measurement method of open-loop gain was outlined last month – and from this a suitable combination of input transconductance and dominant-pole Miller capacitance can be chosen.

Many of the performance graphs shown here are taken from a model (small-signal stages only) amplifier with a Class-A emitter-follower output, at +16 dBu on ±15 V rails. However, since the output from the input pair is in current form, the rail voltage in itself has no significant effect on the linearity of the input stage. It is the current swing at its output that is the crucial factor.

## Vive la differential

The primary motivation for using a differential pair as the input stage of an amplifier is usually its low DC offset. Apart from its inherently lower offset due to the cancellation of the $V_{be}$ voltages, it has the added advantage that its standing current does not have to flow through the feedback network. However a second powerful reason is that its linearity is far superior to single-transistor input stages. Figure 1 shows three versions, in increasing order of sophistication. The resistor-tail version in Figure 1(a) has poor CMRR and PSRR and is generally a false economy; it will not be further considered. The mirrored version in Figure 1(c) has the best balance, as well as twice the transconductance of that in Figure 1(b).

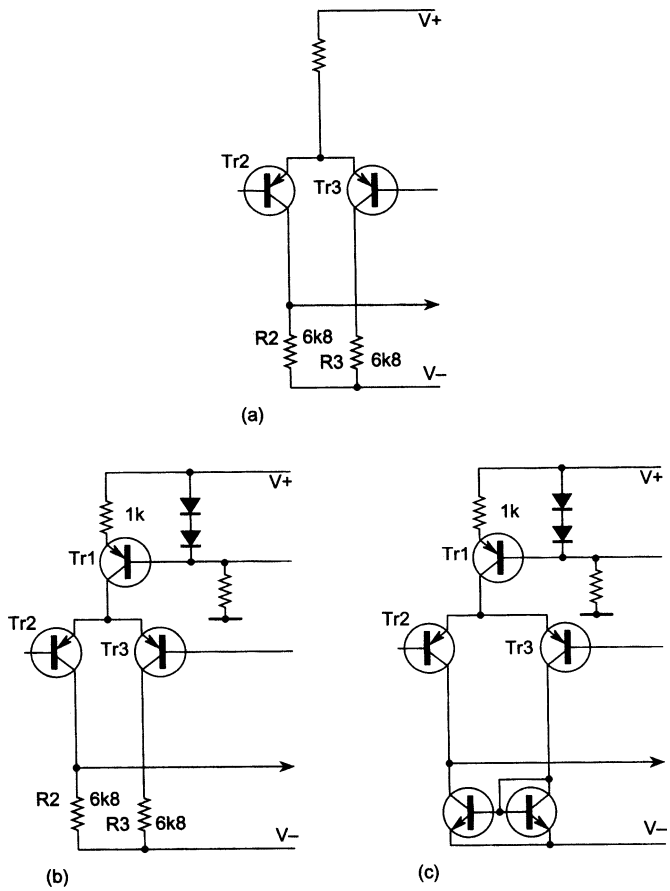



**Figure 1** *Three versions of an input pair: (a) Simple tail resistor; (b) Tail current-source; (c) With collector current-mirror to give inherently good* $I_c$ *balance.*

Intuitively, the input stage should generate a minimal proportion of the overall distortion because the voltage signals it handles are very small, appearing as they do upstream of the VAS that provides almost all the voltage gain. However, above the first pole frequency P1, the current required to drive $C_{\text{dom}}$ dominates the proceedings, and this remorselessly doubles with each octave, thus:

$$I_{\text{pk}} = 2\pi F \cdot C_{\text{dom}} \cdot V_{\text{pk}} \tag{1}$$

For example the current required at 100 W, 8 Ω and 20 kHz, with a 100 pF $C_{\text{dom}}$ is 0.5 mA peak, which may be a large proportion of the input standing current, and so the linearity of transconductance for large current excursions will be of the first importance if we want low distortion at high frequencies.

Figure 2, *curve A*, shows the distortion plot for a model amplifier (at +16 dBu output) designed so that the distortion from all other sources is negligible compared with that from the carefully balanced input stage. With a small-signal class A stage this essentially reduces to making sure that the VAS is properly linearised. Plots are shown for both 80 kHz and 500 kHz measurement bandwidths to show both HF behaviour and LF distortion. It demonstrates that the distortion is below the noise floor until 10 kHz, when it emerges and heaves upwards at a precipitous 18 dB/octave.

This rapid increase is due to the input stage signal current doubling with every octave to drive $C_{\text{dom}}$; this means that the associated third harmonic

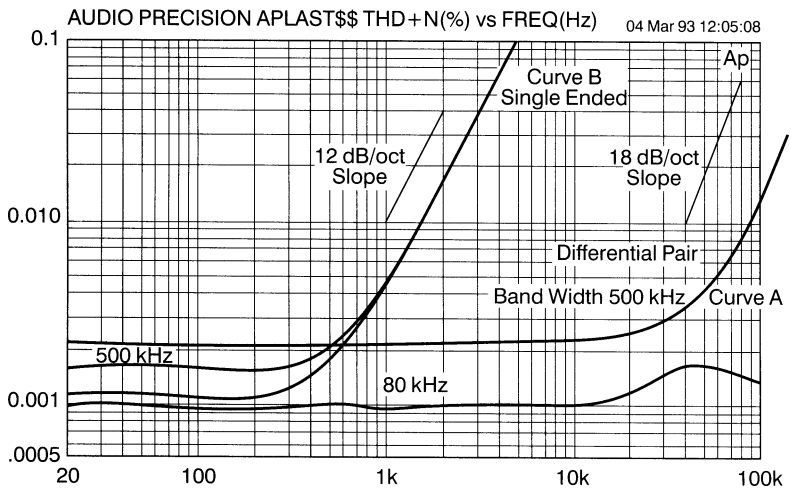



**Figure 2** *Distortion performance of model amplifier differential pair at A compared with singleton input at B. The singleton generates copious second-harmonic distortion.*

distortion will quadruple with every octave increase. Simultaneously the overall NFB available to linearise this distortion is falling at 6 dB/octave since we are almost certainly above the dominant pole frequency P1. The combined effect is an 18 dB/octave rise. If the VAS or the output stage were generating distortion, this would be rising at only 6 dB/octave and would look quite different on the plot.

This form of non-linearity, which depends on the rate-of-change of the output voltage, is the nearest thing to what we normally call TID, an acronym that now seems to be falling out of fashion. Slew-induced-distortion SID is a better description of the effect.

If the input pair is *not* accurately balanced, then the situation is more complex. Second as well as third harmonic distortion is now generated, and by the same reasoning this has a slope of closer to 12 dB/octave. This vital point requires examination.

## Input stage in isolation

The use of a single input transistor (Figure 3 (*a*)) sometimes seems attractive, where the amplifier is capacitor-coupled or has a separate DC servo; it at least promises strict economy. However, the snag is that this singleton configuration has no way to cancel the second-harmonics generated by its strongly-curved exponential $V_{in}/I_{out}$ characteristic.[1] The result is shown in

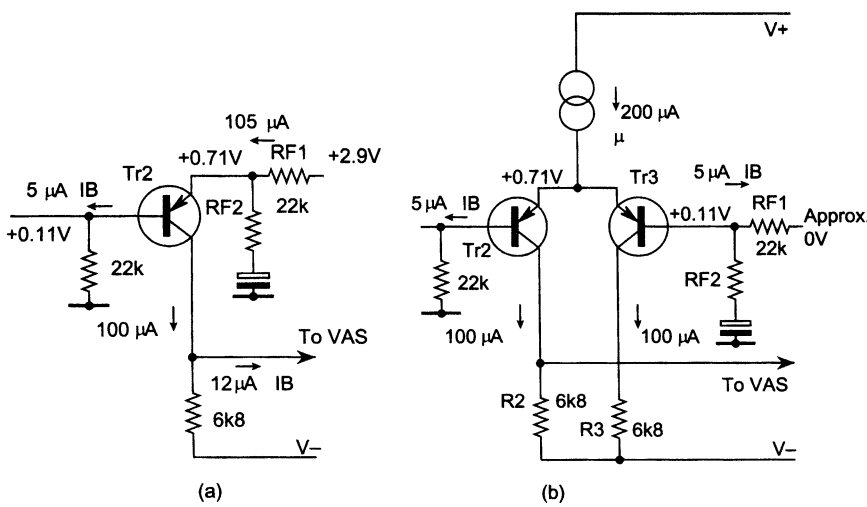



**Figure 3** *Singleton and differential pair input stages showing typical DC conditions. The large DC offset of the singleton* (2.8 V) *is largely due to all the stage current flowing through the feedback resistor RF1.*

Figure 2 curve B, where the distortion is much higher, though rising at the slower rate of 12 dB/octave.

Although the slope of the distortion plot for the whole amplifier tells much, measurement of input-stage nonlinearity in isolation tells more. This may be done with the test circuit of Figure 4. The op-amp uses shunt feedback to generate an appropriate AC virtual earth at the input-pair output. Note that this current-to-voltage conversion op-amp requires a third −30 V rail to allow the i/p pair collectors to work at a realistic DC voltage – i.e. about one diode's-worth above the −15 V rail. $R_f$ can be scaled to stop op-amp clipping without effect to the input stage. The DC balance of the pair may be manipulated by $VR_1$: it is instructive to see the THD residual diminish as balance is approached until, at its minimum amplitude, it is almost pure third harmonic.

The differential pair has the great advantage that its transfer characteristic is mathematically highly predictable.[2] The output current is related to the differential input voltage $V_{in}$ by:

$$I_{out} = I_e \cdot \tanh\left(-V_{in}/2V_t\right) \tag{2}$$

where $V_t$ is the usual 'thermal voltage' of about 26 mV at 25 °C and $I_e$ the tail current.

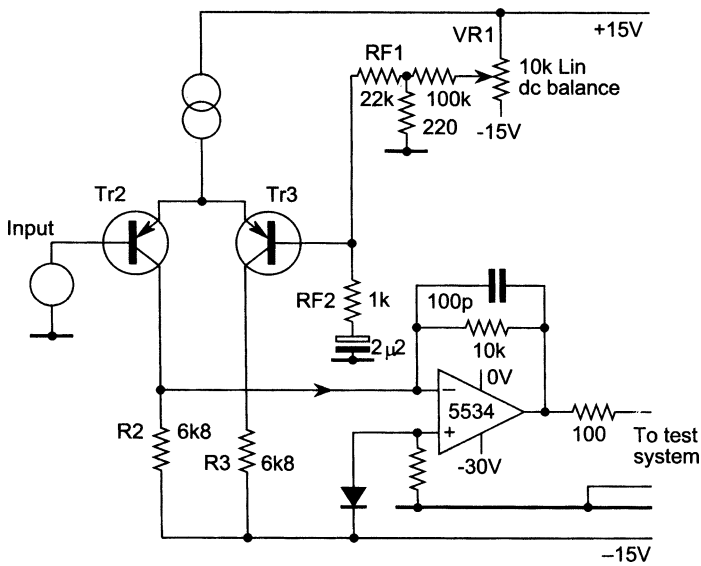



**Figure 4** *Test circuit for examining input stage distortion in isolation. The shunt-feedback opamp is biased to provide the right DC conditions for Tr2.*

This equation demonstrates that the transconductance, $g_m$, is highest at $V_{in} = 0$ when the two collector currents are equal, and that that the value of this maximum is proportional to the tail current, $I_e$. Note also that beta does not figure in the equation, and that the performance of the input pair is not significantly affected by transistor type.

Figure 5(a) shows the linearising effect of local feedback or degeneration on the voltage-in/current-out law. Figure 5(b) plots transconductance against input voltage and demonstrates a reduced peak transconductance

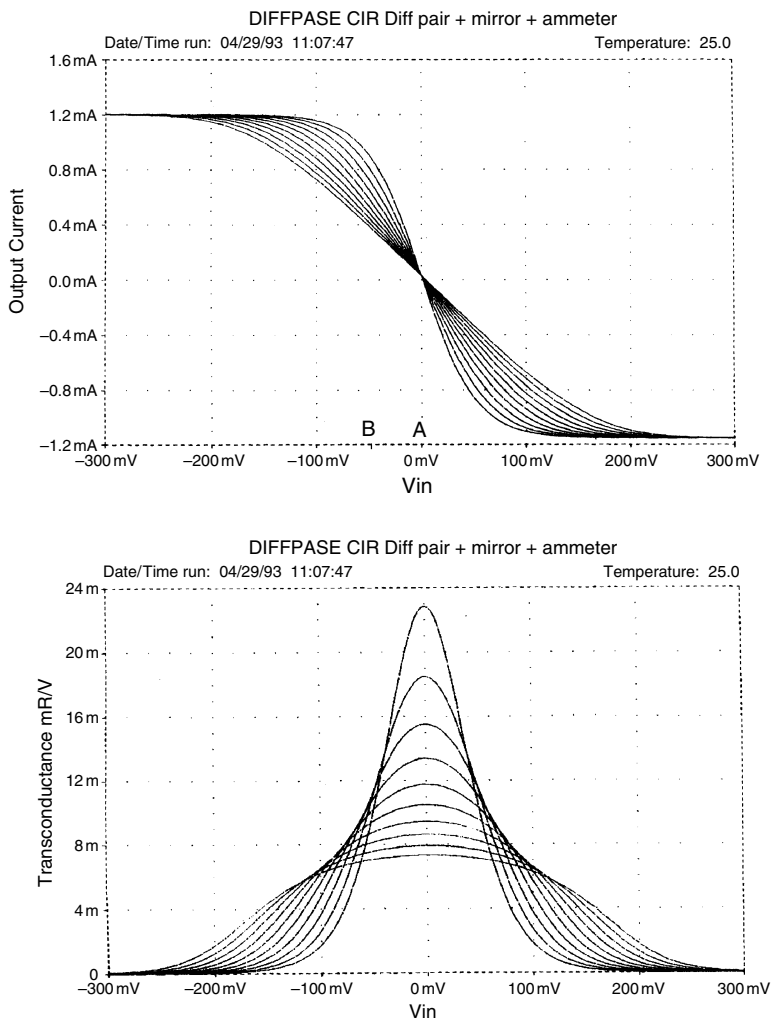



**Figure 5** *Effect of degeneration on input pair V/I law, showing how transconductance is sacrificed in favour of linearity (SPICE simulation).*

value but with the curve made flatter and more linear over a wider operating range. Adding emitter degeneration markedly improves input stage linearity at the expense of noise performance. Overall amplifier feedback factor is also reduced since the HF closed-loop gain is determined solely by the input transconductance and the value of the dominant-pole capacitor.

## Input stage balance

One relatively unknown property of the differential pair in power amplifiers is its sensitivity to exact DC balance. Minor deviations from equality of $I_c$ in the pair seriously upset the second-harmonic cancellation by moving the operating point from $A$ in Figure 5(a) to $B$. Since the average slope of the characteristic is greatest at $A$, serious imbalance also reduces the open-loop gain. The effect of small amounts of imbalance is shown in Figure 6 and Table 1: for an input of −45 dBu a collector current imbalance of only 2% increases THD from 0.10% to 0.16%; for 10% imbalance this deteriorates to 0.55%. Unsurprisingly, imbalance in the other direction ($I_{c1} > I_{c2}$) gives similar results.

This gives insight[4] into the complex changes that accompany the simple changing the value of $R_2$. For example, we might design an input stage as per Figure 7(a), where $R_1$ has been selected as 1 kΩ by uninspired

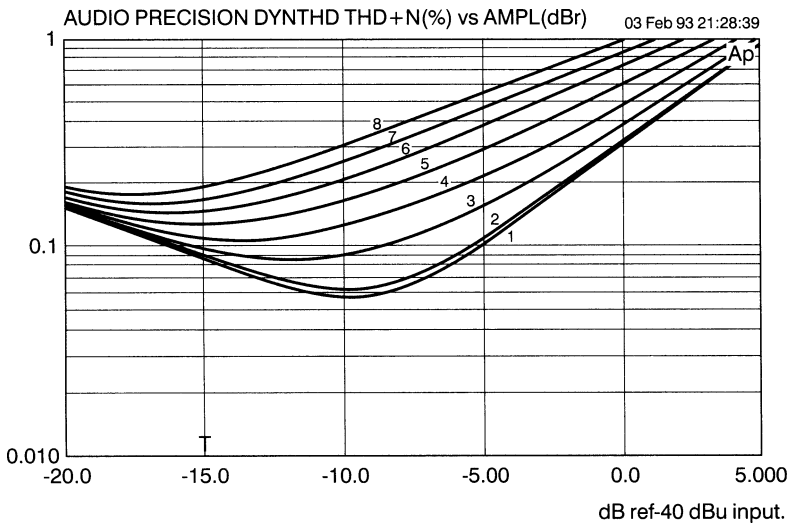



**Figure 6** *Effect of collector-current imbalance on an isolated input pair; the second harmonic rises well above the level of the third if the pair moves away from balance by as little as* 2%.

**Table 1** *Key to Figure 6*

| *Curve No.* | $I_c$ *Imbalance* (%) |
|---|---|
| 1 | 0 |
| 2 | 0.5 |
| 3 | 2.2 |
| 4 | 3.6 |
| 5 | 5.4 |
| 6 | 6.9 |
| 7 | 8.5 |
| 8 | 10 |

Imbalance defined as deviation of $I_c$ (per device) from that value which gives equal currents in the pair.

guesswork and $R_2$ made highish at 10 kΩ in a plausible but misguided attempt to maximise o/l gain by minimising loading on $Tr_1$ collector. $R_3$ is also made 10 kΩ to give the stage a notional 'balance', though unhappily this is a visual rather than electrical balance. The asymmetry is shown in the resulting collector currents: this design will generate avoidable second harmonic distortion, displayed in the 10 kΩ curve of Figure 8.

However, recognising the importance of DC balancing, the circuit can be rethought as per Figure 7(b). If the collector currents are to be roughly balanced, then $R_2$ must be about $2 \times R_1$, as both have about 0.6 V across them. The effect of this change is shown in the 2.2 kΩ curve of Figure 8. The improvement is accentuated as the o/l gain has also increased by some 7 dB, though this has only a minor effect on the closed-loop linearity compared with the improved balance of the input pair. $R_3$ has been excised as it contributes little to stage balance.

## The joy of current mirrors

While the input pair can be approximately balanced by the correct choice of $R_1$ and $R_2$, other circuit tolerances are significant and Figure 6 shows that balance is critical, needing to be accurate to at least 1% for optimal linearity. The standard current-mirror configuration shown in Figure 7(c) forces the two collector currents very close to equality, giving proper cancellation of second harmonic. The resulting improvement shows up in the current-mirror curve of Figure 8. There is also less DC offset due to unequal base currents flowing through input and feedback resistances; we often find that a power-amplifier improvement gives at least two separate benefits. This simple mirror has its own residual base current errors but they are not large enough to affect distortion.

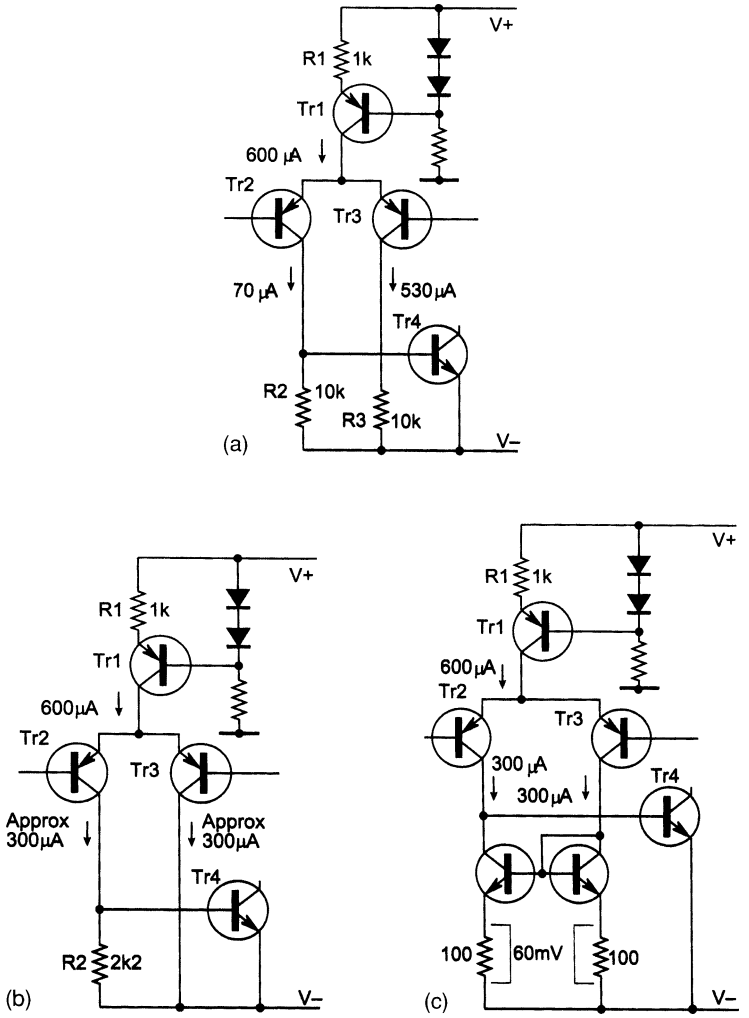



**Figure 7** *Improvements to the input pair: (a) Poorly designed version; (b) Better . . . partial balance by correct choice of* $R_2$. *(c) Best . . . near-perfect Ic balance enforced by mirror.*

The hyperbolic tangent law also holds for the mirrored pair,[3] though the output current swing is twice as great for the same input voltage as the resistor-loaded version. This doubled output occurs at the same distortion level as for the single-ended version, as linearity depends on the input voltage, which has not changed. Alternatively, to get the same output we can halve the input which, with a properly balanced pair generating only third harmonic, will produce just one-quarter the distortion, a pleasing result.

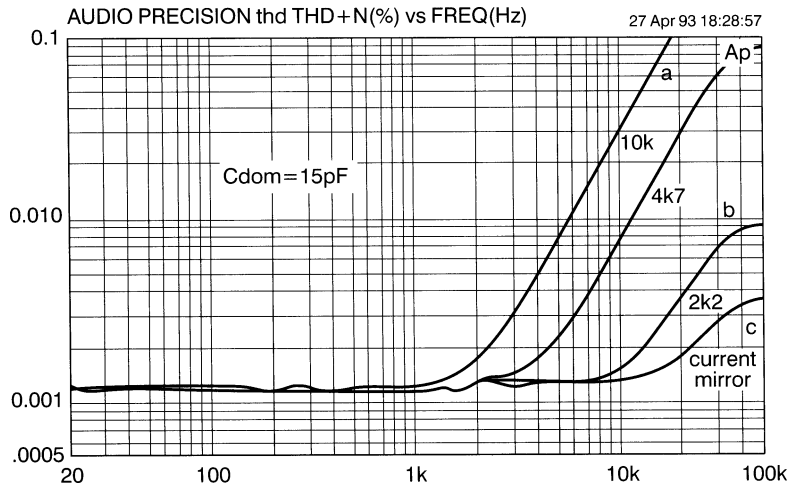



**Figure 8** *Distortion of model amplifier: (a) Unbalanced with* $R_2 = 10\,k\Omega$*; (b) Partially balanced with* $R = 2.2\,k\Omega$*; (c) Accurately balanced by current-mirror.*

A low cost mirror made from discrete transistors forgoes the $V_{be}$ matching available to IC designers, and so requires its own emitter degeneration for good current-matching. A voltage drop across the mirror emitter resistors in the range 30–60 mV will be enough to make the effect of $V_{be}$ tolerances on distortion negligible If degeneration is omitted, there is significant variation in HF distortion performance with different specimens of the same transistor type. Adding a current mirror to a reasonably well balanced input stage will increase the total o/l gain by at least 6 dB, and by up to 15 dB if the stage was previously poorly balanced. This needs to be taken into account in setting the compensation. Another happy consequence is that the slew-rate will be roughly doubled, as the input stage can now source and sink current into $C_{dom}$ without wasting it in a collector load. If $C_{dom}$ is 100 pF, the slewrate of Figure 7(b) is about 2.8 V/μs up and down, while Figure 7(c) gives 5.6 V/μs. The unbalanced pair in Figure 7(a) displays further vices by giving 0.7 V/μs positive-going and 5 V/μs negative-going.

## Improving linearity

Now that the input pair has been fitted with a mirror, we may still feel that the HF distortion needs further reduction; after all, once it emerges from the noise floor it goes up eight times with each doubling of frequency, and so it is well worth pushing the turn point as far as possible up the

frequency range. The input pair shown has a conventional value of tail-current. We have seen that the stage transconductance increases with $I_c$, and so it is possible to increase the $g_m$ by increasing the tail-current, and then return it to its previous value (otherwise $C_{dom}$ would have to be increased proportionately to maintain stability margins) by applying local NFB in the form of emitter-degeneration resistors. This ruse powerfully improves input linearity despite its rather unsettling flavour of something-for-nothing. The transistor nonlinearity can here be regarded as an internal nonlinear emitter resistance $r_e$, and what we have done is to reduce the value of this (by increasing $I_c$) and replace the missing part of it with a linear external resistor, $R_e$.

For a single device, the value of $r_e$ can be approximated by:

$$r_e = 25/I_c\,\Omega \text{ (for } I_c \text{ in mA)}. \tag{3}$$

Our original stage at Figure 9(a) has a perdevice $I_c$ of 600 µA, giving a differential (i.e. mirrored) $g_m$ of 23 mA/V and $r_e = 41.6\,\Omega$. The improved version at Figure 9(b) has $I_c = 1.35$ mA and so $r_e = 18.6\,\Omega$. Emitter degeneration resistors of 22 Ω are required to reduce the $g_m$ back to its original value, as $18.6 + 22 = 41.6$. The distortion measured by the circuit of Figure 4 for a −40 dBu input voltage is reduced from 0.32% to 0.032%, which is an extremely valuable linearisation, and will translate into a distortion reduction at HF of about five times for a complete amplifier. For reasons that will emerge later the full advantage is rarely gained. The distortion remains a visually pure third harmonic so long as the input pair

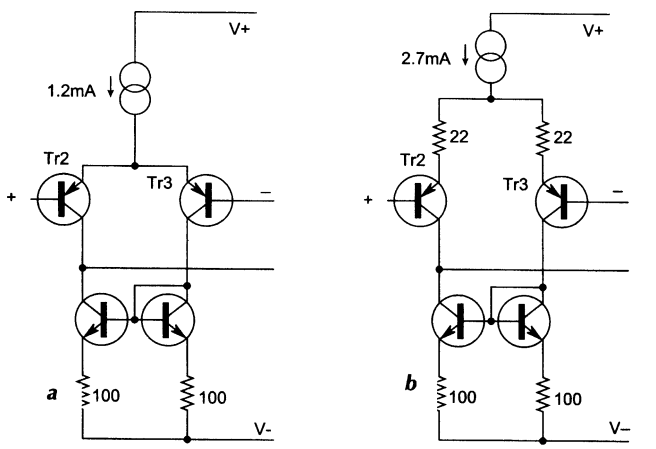



**Figure 9** *Input pairs before and after constant-$g_m$ degeneration showing how to double stage current while keeping transconductance constant: distortion is reduced by about ten times.*

remains balanced. Clearly this sort of thing can only be pushed so far, as the reciprocal-law reduction of $r_e$ is limited by practical values of tail current. A name for this technique seems to be lacking; 'constant-$g_m$ degeneration' is descriptive but rather a mouthful.

Since the standing current is roughly doubled so has the slew rate: 10 V/μs to 20 V/μs. Once again we gain two benefits for the price of one modification.

For still better linearity, various techniques exist. When circuit linearity needs a lift, it is often a good approach to increase the *local* feedback factor, because if this operates in a tight local NFB loop there is often little effect on the overall global-loop stability. A reliable method is to replace the input transistors with complementary-feedback (CFP or Sziklai) pairs, as shown in the stage of Figure 10(a). If an isolated input stage is measured using the test circuit of Figure 4, the constant $g_m$ degenerated version shown in Figure 9(b) yields 0.35% third-harmonic distortion for a −30 dBu input voltage, while the CFP version gives 0.045%. Note that the input level here is 10 dB up on the previous example to get well clear of the noise floor. When this stage is put to work in a model amplifier, the third-harmonic distortion at a given frequency is roughly halved, assuming other distortion sources have been appropriately minimised. However, given the steep slope of input stage distortion, this extends the low distortion regime up in frequency by less than an octave. See Figure 11.

The CFP circuit does require a compromise on the value of $R_e$, which sets the proportion of the standing current that goes through the NPN and PNP devices on each side of the stage. In general, a higher value of $R_e$ gives better linearity, but more noise, due to the lower $I_c$ in the NPN devices that are the inputs of the input stage, as it were, causing them to match less well the relatively low source resistances. 2.2 kΩ is a reasonable compromise.

Other elaborations of the basic input pair are possible. Power amp design can live with a restricted common-mode range in the input stage that would be unusable in an op-amp, and this gives the designer great scope. Complexity in itself is not a serious disadvantage as the small-signal stages of the typical amplifier are of almost negligible cost compared with mains transformers, heatsinks, etc.

Two established methods to produce a linear input transconductance stage (often referred to in opamp literature simply as a transconductor) are the cross-quad[5] and the cascomp[6] configurations. The cross-quad (Figure 10(b)) gives a useful reduction in input distortion when operated in isolation but is hard to incorporate in a practical amplifier because it relies on very low source resistance to tame the negative conductances inherent in its operation. The cross-quad works by imposing the input voltage to each half across two base-emitter junctions in series, one in each arm of the circuit. In theory the errors due to non-linear $r_e$ of the transistors is divided by beta, but in practice things seem less rosy.

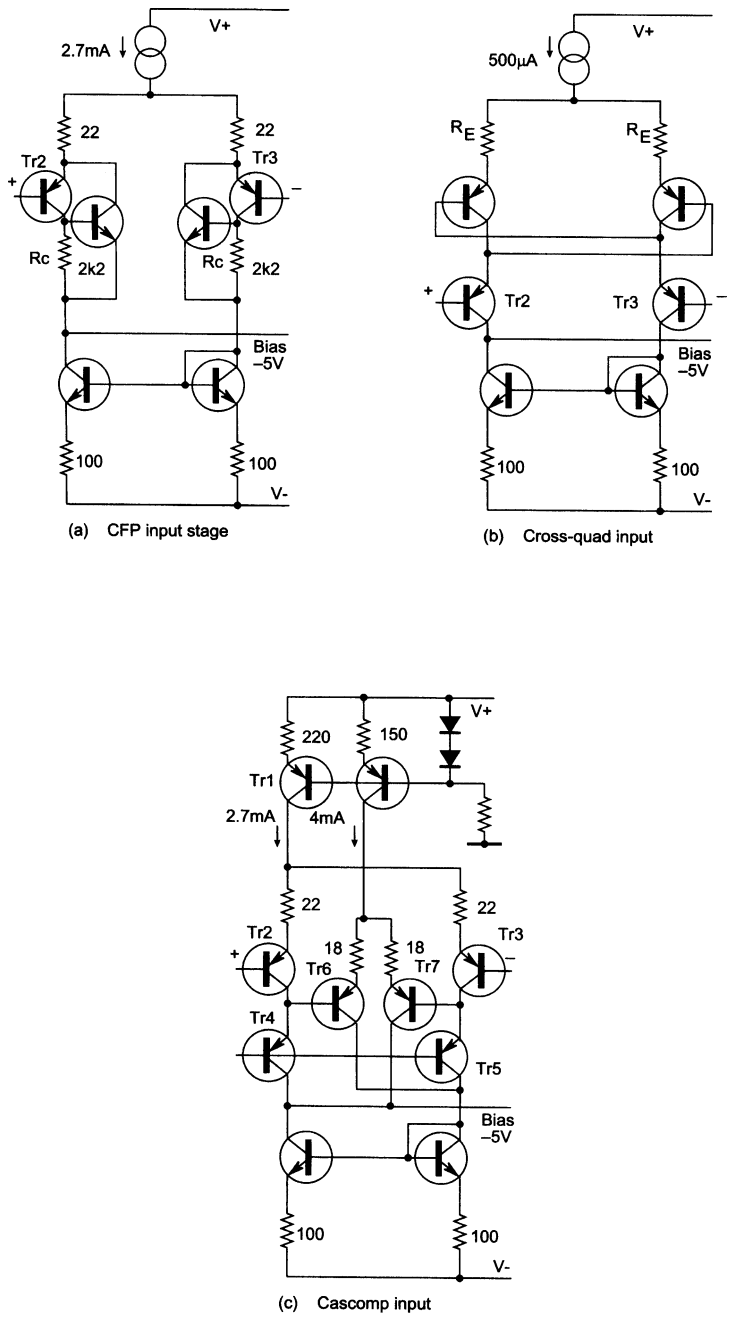



**Figure 10** *Some enhanced differential pairs: (a) The complementary feedback pair; (b) The cross-quad; (c) The cascomp.*

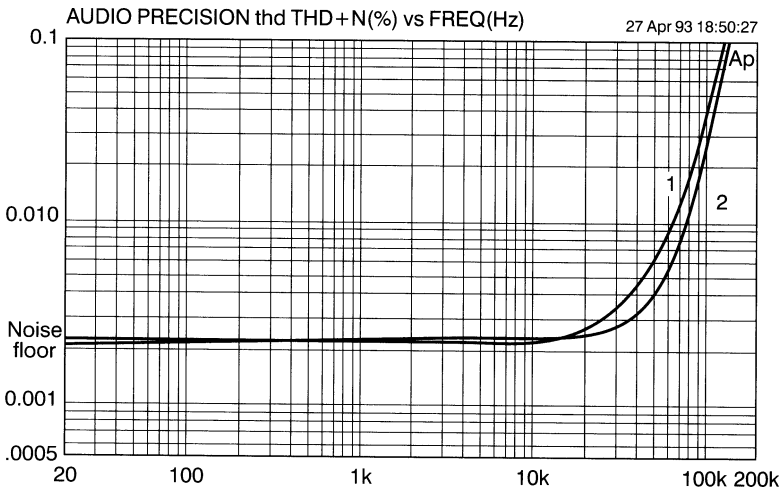



**Figure 11** *Whole-amplifier THD with normal and CFP input stages; input stage distortion only shows above noise floor at* 20 kHz*, so improvement occurs above this frequency. The noise floor appears high as the measurement bandwidth is* 500 kHz.

The cascomp (Figure 10(c)) does not have this snag, though it is significantly more complex to design. $Tr_2$, $Tr_3$ are the main input pair as before, delivering current through cascode transistors $Tr_4$, $Tr_5$ (this does not in itself affect linearity) which, since they carry almost the same current as $Tr_2$, $Tr_3$ duplicate the input $V_{bev}$ errors at their emitters. This is sensed by error diff-amp $Tr_6$, $Tr_7$ whose output currents are summed with the main output in the correct phase for error-correction. By careful optimisation of the (many) circuit variables, distortion at −30 dBu input can be reduced to about 0.016% with the circuit values shown. Sadly, this effort provides very little further improvement in whole-amplifier HF distortion over the simpler CFP input, as other distortion mechanisms are coming into play – for instance the finite ability of the VAS to source current into the other end of $C_{dom}$.

Power amplifiers with pretensions to sophistication sometimes add cascoding to the standard input differential amplifier. This does nothing to improve input stage linearity as there is no appreciable voltage swing on the input collectors; its main advantage is reduction of the high $V_{ce}$ that the input devices work at. This allows cooler running, and therefore possibly improved thermal balance; a $V_{ce}$ of 5 V usually works well. Isolating the input collector capacitance from the VAS input often allows $C_{dom}$ to be somewhat reduced for the same stability margins, but it is doubtful if the advantages really outweigh the increased complexity.

## Other considerations

As might be expected, the noise performance of a power amplifier is set by the input stage, and so it is briefly examined here. Power amp noise is not an irrelevance: a powerful amplifier is bound to have a reasonably high voltage gain and this can easily result in a faint but irritating hiss from efficient loudspeakers even when the volume control is fully retarded. In the design being evolved here the EIN has been measured at −120 dBu, which is only 7 or 8 dB inferior to a first-class microphone preamplifier. The inferiority is largely due to the source resistances seen by the input devices being higher than the usual 150 Ω microphone impedance. For example, halving the impedance of the feedback network shown in ptl (22 kΩ and 1 kΩ) reduces the EIN by approx 2 dB.

Slew rate is another parameter usually set by the input stage, and has a close association with HF distortion. The amplifier slew rate is proportional to the input's maximum-current capability, most circuit configurations being limited to switching the whole of the tail current to one side or the other. The usual differential pair can only manage half of this, as with the output slewing negatively half the tail-current is wasted in the input collector load $R_2$. The addition of an input current-mirror, as advocated, will double the slew rate in both directions. With a tail current of 1.2 mA, the slew rate is improved from about 5 V/μs to 10 V/μs. (for $C_{dom} = 100$ pF) The constant $g_m$ degeneration method of linearity enhancement in Figure 9 further increases it to 20 V/μs. The mathematics of voltage-slewing is simple:

Slew rate $= I/C_{dom}$ in V/μs for maximum I in μA, $C_{dom}$ in pF.

The maximum output frequency for a given slew rate and voltage is:

$$F_{max} = S_r/2\pi V_{pk} = S_r/(2\pi \cdot \sqrt{2} \cdot V_{rms}) \tag{4}$$

Likewise, a sinewave of given amplitude has a maximum slew-rate (at zero-crossing) of:

$$S_{r\,max} = \mathrm{d}V/\mathrm{d}t = \omega_{max} \cdot V_{pk} = 2\pi F V_{pk} \tag{5}$$

So, for example, with a slew rate of 20 V/μs the maximum frequency at which 35 V r.m.s. can be sustained is 64 kHz, and if $C_{dom}$ is 100 pF, then the input stage must be able to source and sink 2 mA peak.

A vital point is that the current flowing through $C_{dom}$ must be sourced/sunk by the VAS as well as the input pair. Sinking is usually no problem, as the VAS common-emitter transistor can be turned on as hard as required. The current source or bootstrap at the VAS collector will however have a limited sourcing ability, and this can often turn out to be an unexpected limitation on the positive-going slew rate.

## References

1. Gray and Meyer, *Analysis & Design of Analog Integrated Circuits*, Wiley 1984, p 172 (*exponential law of singleton*).
2. Gray and Meyer, *Analysis and Design of Analog Integrated Circuits,* p 194 (*tanh law of simple pair*).
3. Gray and Meyer, *Analysis and Design of Analog Integrated Circuits,* p 256 (*tanh law of current-mirror pair*).
4. Self, *Sound Mosfet Design Electronics & Wireless World*, September 1990.
5. Feucht, *Handbook of Analog Circuit Design*, Academic Press 1990, p 432.
6. Quinn, *IEEE International Solid-State Circuits Conference*, THPM 14.5, p 188 (Cascomp).

# 18 Distortion in power amplifiers, Part III: the voltage-amplifier stage

*October 1993*

The voltage-amplifier stage (or VAS) has often been regarded as the most critical part of a power-amplifier, since it not only provides all the voltage gain but also must deliver the full output voltage swing. This is in contrast to the input stage which may give substantial transconductance gain, but the output is in the form of a current. But as is common in audio design, all is not quite as it appears. A well-designed voltage amplifier stage will contribute relatively little to the overall distortion total of an amplifier, and if even the simplest steps are taken to linearise it further, its contribution sinks out of sight.

As a starting point, Figure 1 shows the distortion plot of a model amplifier with a Class-A output (±15 V rails, +16 dBu out). The model is as described in previous articles. No special precautions have been taken to linearise the input stage or the VAS and output stage distortion is negligible. It can be seen that the distortion is below the noise floor at low frequencies; the distortion slowly rising from about 1 kHz is coming from the voltage amplifier stage. At higher frequencies, where the VAS 6 dB/octave rise becomes combined with the 12 or 18 dB/octave rise of input stage distortion, we can see the accelerating distortion slope typical of many amplifier designs.

The main reason why the voltage amplifier stage generates relatively little distortion is because at LF, global feedback linearises the whole amplifier, while at HF the voltage amplifier stage is linearised by local negative feedback through $C_{\mathrm{dom}}$.

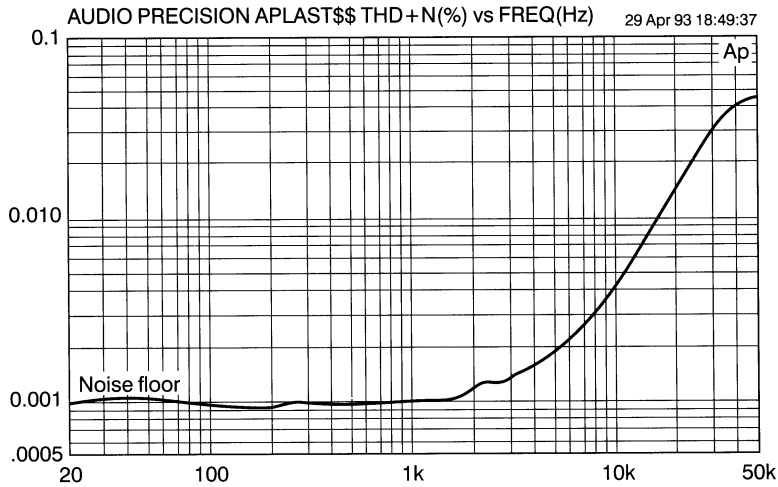



**Figure 1** *THD plot for model amp showing distortion below noise floor at low frequency, and increasing from* 2 kHz *to* 20 kHz. *The ultimate roll-off is due to the* 80 kHz *measurement bandwidth.*

## Examining the mechanism

Isolating the voltage amplifier stage distortion for study requires the input pair to be specially linearised, or else its steeply rising distortion characteristic will swamp the VAS contribution. This is most easily done by degenerating the input stage which also reduces the open-loop gain. The reduced feedback factor mercilessly exposes voltage amplifier stage nonlinearity. This is shown in Figure 2, where the 6 dB/octave slope suggests origination in the VAS, and increases with frequency solely because the compensation is rolling-off the global feedback factor.

Confirming that this distortion is due solely to the voltage amplifier stage requires varying VAS linearity experimentally while leaving other circuit parameters unchanged. Figure 3 achieves this by varying the VAS negative rail voltage; this varies the proportion of its characteristic over which the voltage amplifier stage swings, and thus only alters the effective VAS linearity, as the important input stage conditions remain unchanged. The current-mirror must go up and down with the VAS emitter for correct operation, and so the $V_{ce}$ of the input devices also varies, but this has no significant effect as can be proved by the unchanged behaviour on inserting cascode stages in the input transistor collectors.

The typical topology as shown in Figure 4(a) is a classical common emitter voltage amplifier stage with a current-drive input into the base. The small-signal characteristics, which set open-loop gain and so on, can be usefully simulated by the spice model shown in Figure 5, of a VAS

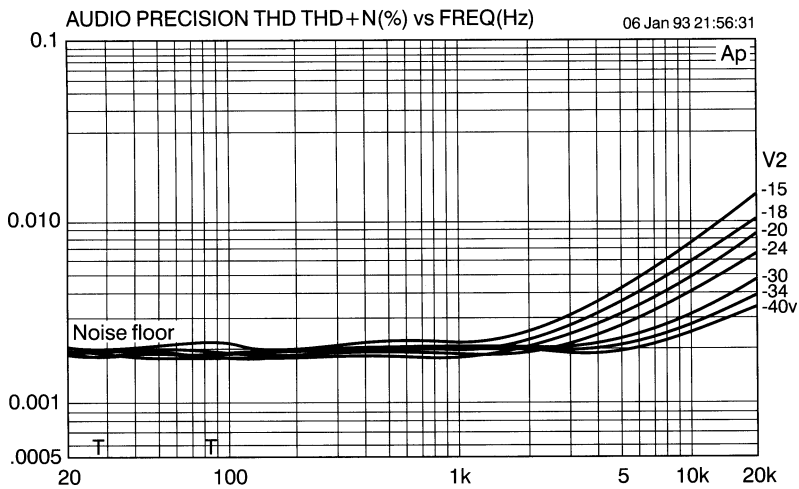



**Figure 2** *The change in HF distortion resulting from varying the negative rail in the VAS test circuit. The voltage amplifier stage distortion is only revealed by degenerating the input stage with* 100 Ω *resistors.*

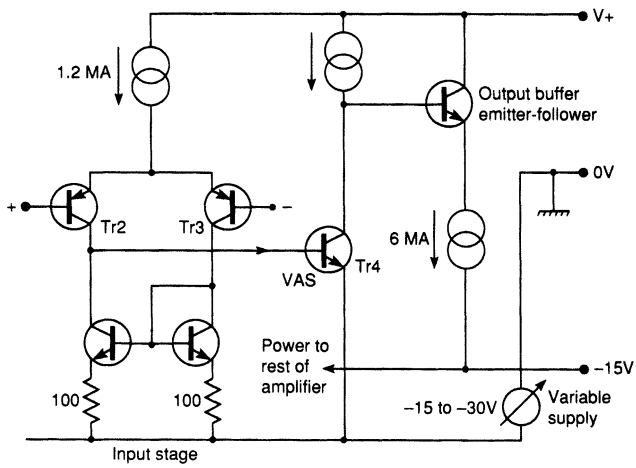



**Figure 3** *Voltage amplifier stage distortion test circuit. Although the input pair mirror moves up and down with the VAS emitter, the only significant parameter being varied is the available voltage swing at the collector.*

reduced to its conceptual essentials. $G$ is a current source whose value is controlled by the voltage-difference between $R_{\text{in}}$ and $R_{f2}$, and represents the differential transconductance input stage. $F$ represents the voltage amplifier stage transistor, and is a current source yielding a current of beta

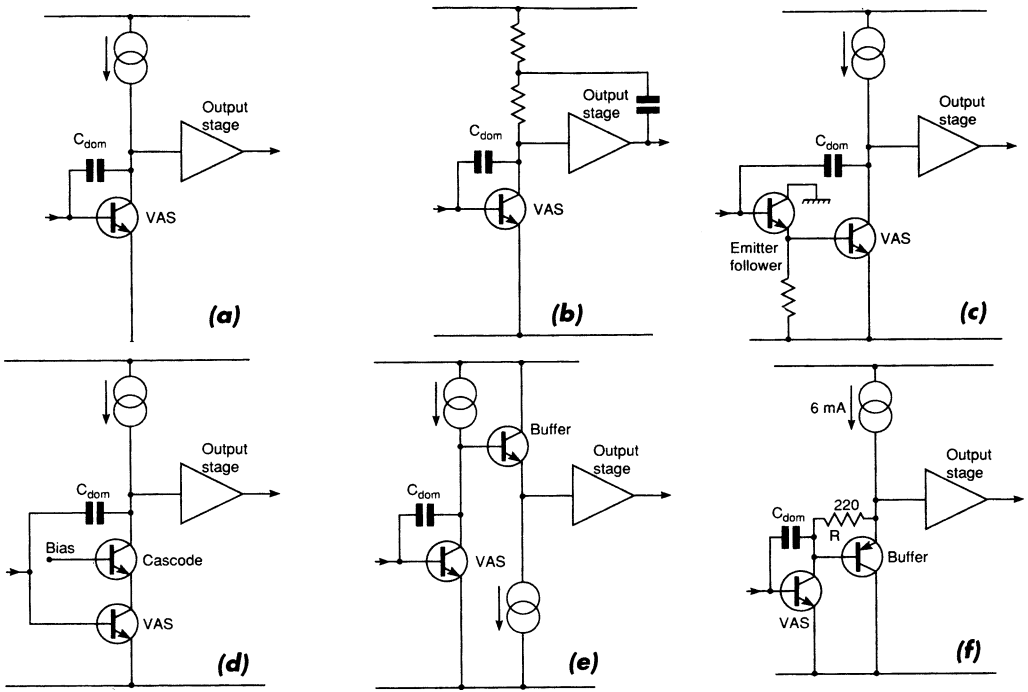



**Figure 4** *Six variations on a voltage amplifier stage: (a) conventional current source VAS, (b) conventional bootstrapped VAS, (c) increase in local NFB by adding emitter follower, (d) increase in local NFB by cascoding, (e) one method of buffering VAS collector from output stage, (f) alternative buffering arrangement uses bootstrapping resistor.*

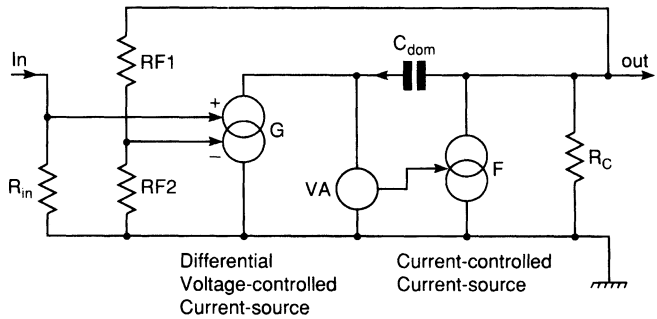



**Figure 5** *Conceptual spice model of differential input stage* (G) *and VAS* (F). *The current in* F *is beta times the current in* G.

times that sensed flowing through ammeter $V$ which, by spice convention, is a voltage source set to 0 V.

The value of beta, representing current-gain, models the relationship between VAS collector current and base current. $R_c$ represents the total stage collector impedance, a typical real value being 22 kΩ. With suitable parameter values, this simple model provides a useful demonstration of relationships between gain, dominant-pole frequency, and input stage current outlined in the first article in this series. Injecting a small signal current into the output node from an extra current source also allows the fall of impedance with frequency to be examined.

The overall voltage gain clearly depends linearly on beta, which in real transistors may vary widely. Working on the trusty engineering principle that what cannot be controlled must be made irrelevant, local shunt NFB through $C_{dom}$ sets the crucial HF gain that controls Nyquist stability. The LF gain below the dominant pole frequency P1 remains variable (and therefore so does P1) but is ultimately of little importance; if there is an adequate NFB factor for overall linearisation at HF then there are unlikely to be problems at LF where gain is highest. As for the input stage, the linearity of the voltage amplifier stage is not greatly affected by transistor type, given a reasonably high beta value.

## Stage distortion

Voltage amplifier stage distortion arises from a curved transfer characteristic of the common-emitter amplifier, a small portion of an exponential.[1] This characteristic generates predominantly second-harmonic distortion, which, in a closed-loop amplifier, will increase at 6 dB/octave with frequency.

Distortion does not get worse for more powerful amplifiers as the stage traverses a constant proportion of its characteristic as the supply-rails are

increased. This is not true of the input stage: increasing output swing increases the demands on the transconductance amp as the current to drive $C_{dom}$ increases. The increased $V_{ce}$ of the input devices does not measurably affect their linearity.

It seems ironic that VAS distortion only becomes clearly visible when the input pair is excessively degenerated – a pious intention to 'linearise before applying feedback' can make the closed loop distortion worse by reducing the open loop gain and hence the NFB factor available to linearise the VAS. In a real (non-model) amplifier with a distortive output stage, the deterioration will be worse.

The local open-loop gain of the VAS (that existing inside the local feedback loop closed by $C_{dom}$) should be high, so that the voltage amplifier stage can be linearised. This precludes a simple resistive load. Increasing the value of $R_c$ will decrease the collector current of the transistor reducing its transconductance. This reduces voltage gain to the starting value.

One way to ensure sufficient gain is to use an active load. Either bootstrapping or a current source will do this effectively, though the current source is perhaps more dependable and is the usual choice for hi-fi or professional amplifiers.

The bootstrap promises more output swing as the collector of $Tr_4$ can soar above the positive rail. This suits applications such as automotive power amps that must make the best possible use of a restricted supply voltage.[2]

These two active-load techniques also ensure enough current to drive the upper half of the output stage in a positive direction right up to the supply rail. If the collector load were a simple resistor, this capability would certainly be lacking.

Checking the effectiveness of these measures is straightforward. The collector impedance may be determined by shunting the collector node to ground with decreasing resistance until the open loop gain falls by 6 dB indicating that the collector impedance is equal to the current value of the test resistor.

The popular current source version is shown in Figure 4(a). This works well, though the collector impedance is limited by the effective output resistance $R_o$ of the voltage amplifier stage and the current source transistors[3] which is another way of saying that the improvement is limited by Early effect.

It is often stated that this topology provides current-drive to the output stage; this is only partly true. It is important to realise that once the local NFB loop has been closed by adding $C_{dom}$ the impedance at the VAS output falls at 6 dB/octave for frequencies above P1. The impedance is only a few kΩ at 10 kHz, and this hardly qualifies as current-drive at all.

Bootstrapping (Figure 4(b)) works in most respects as well as a current source load, for all its old-fashioned flavour. The method has been

criticised for prolonging recovery from clipping. I have no evidence to offer on this myself, but I can state that a subtle drawback definitely exists: LF open loop gain is dependent on amplifier output loading. The effectiveness of bootstrapping depends crucially on the output stage gain being unity or very close to it. However the presence of the output transistor emitter resistors means that there will be a load-dependant gain loss in the output stage significantly altering the amount by which the VAS collector impedance is increased. Hence the LF feedback factor is dynamically altered by the impedance characteristics of the loudspeaker load and the spectral distribution of the source material.

This has significance if the load is a quality speaker with impedance modulus down to 2 Ω, in which case the gain loss is serious. If anyone needs a new audio-impairment mechanism to fret about, then I offer this one in the confident belief that its effects, while measurable, are not of audible significance.

The standing d.c. current also varies with rail voltage. Since accurate setting and maintaining of quiescent current is difficult enough, an extra source of possible variation is decidedly unwelcome.

A less well known but more dependable form of bootstrapping is available if the amplifier incorporates a unity gain buffer between the VAS collector and the output stage as shown in Figure 4(f), where $R_c$ is the collector load, defining the VAS collector current by establishing the $V_{be}$ of the buffer transistor across itself. This is constant, and $R_c$ is therefore bootstrapped and appears to the VAS collector as a constant current source.

In this sort of topology a voltage amplifier stage current of 3 mA is quite sufficient, compared with the 6 mA standing current in the buffer stage. The voltage amplifier stage would in fact work well with collector currents down to 1 mA, but this tends to compromise linearity at the high-frequency, high-voltage corner of the operating envelope, as the VAS collector current is the only source for driving current into $C_{dom}$.

## Voltage stage enhancements

Figure 2, which shows only VAS distortion, clearly indicates the need for further improvement over that given inherently by the presence of $C_{dom}$ if an amplifier is to avoid distortion. While the virtuous approach might be an attempt to straighten the curved voltage amplifier stage characteristic, in practice the simplest method is to increase the amount of local negative feedback through this capacitance. Equation 1 in the first article shows that the LF gain (i.e. the gain before $C_{dom}$ is connected) is the product of input stage transconductance, $Tr_4$ beta and the collector impedance $R_c$. The last two factors represent the VAS gain and therefore the amount of local NFB can be augmented by increasing either. Note that so long as

the value of $C_{dom}$ remains the same, the global feedback factor at HF is unchanged and so stability is not affected.

The effective beta of the stage can be substantially increased by replacing the VAS transistor with a Darlington, Figure 4(c). Adding an extra stage to a feedback amplifier always requires thought because, if significant additional phase-shift is introduced, the global loop stability may suffer. In this case the new stage is inside the Miller loop and so there is little likelihood of trouble. The function of such an emitter follower is sometimes described as 'buffering the input stage from the VAS' but its true function is linearisation by enhancement of local NFB.

Alternatively the stage collector impedance may be increased for higher local gain. This is could be done with a cascode configuration (Figure 4(d)) but the technique is only useful when driving a linear impedance rather than a Class-B output stage with its non-linear input impedance.

Assuming for the moment that this problem is dealt with, either by use of a Class-A output or by VAS-buffering, the drop in distortion is dramatic as is the beta-enhancement method. The gain increase is ultimately limited by Early effect in the cascode and current source transistors, and more seriously by the loading effect of the next stage. But it is of the order of 10 times and gives a useful improvement.

This is shown by curves A, B in Figure 6 where the input stage of a model amplifier has been over-degenerated with 100 Ω emitter resistors to bring out the voltage amplifier stage distortion more clearly.

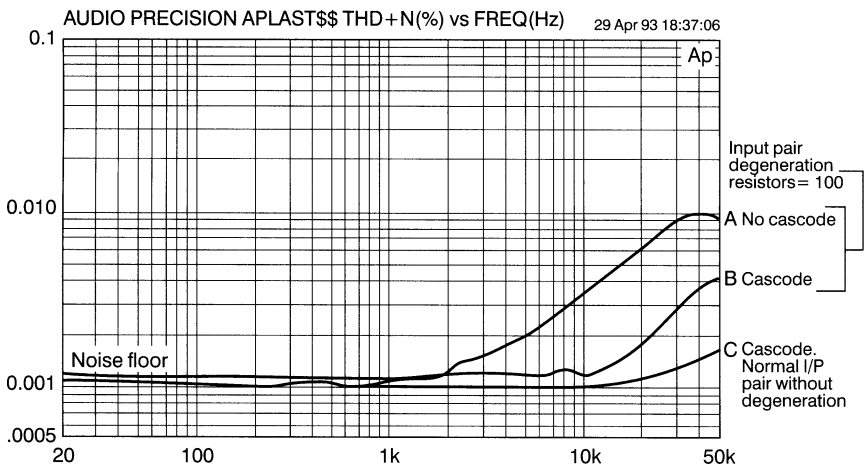



**Figure 6** *Showing the reduction of VAS distortion possible by cascoding. The results from adding an emitter follower to the voltage amplifier stage, as an alternative method of increasing local voltage amplifier stage feedback, are very similar.*

Note that in both cases the slope of the distortion increase is 6 dB/octave. Curve C shows the result when a standard undegenerated input pair is combined with the cascoded VAS; the distortion is submerged in the noise floor for most of the audio band, being well below 0.001%.

This justifies my assertion that input stage and VAS distortion need not be a problem; we have all but eliminated distortions 1 and 2 from the list of seven given in the first article.

A cascode transistor also allows the use of a high-beta transistor for the voltage amplifier stage; these typically have a limited $V_{ceo}$ that cannot withstand the high rail voltages of a high-power amplifier. There is a small loss of available voltage swing, but only about 300 mV, which is usually tolerable. Experiment shows that there is nothing to be gained by cascoding the current source collector load.

A cascode topology is often used to improve frequency response by isolating the upper collector from the $C_{bc}$ of the lower transistor. In this case the frequency response is deliberately defined by a well defined passive component.

It is hard to say which technique is preferable; the emitter follower circuit is slightly simpler than the cascode version, which requires extra bias components, but the cost difference is minimal. When wrestling with these kind of financial decisions it as well to remember that the cost of a small-signal transistor is often less than a fiftieth of that of an output device, and the entire small-signal section of an amplifier usually represents less than 1% of the total cost, when heavy metal such as the mains transformer and heatsinks are included.

## Benefits of voltage drive

The fundamentals of linear voltage amplifier stage operation require that the collector impedance is high, and not subject to external perturbations. Thus a Class-B output stage, with large input impedance variations around the crossover point, is the worst possible load. The 'standard' amplifier configuration deserves tribute that it can handle this internal unpleasantness gracefully, 100 W/8 Ω distortion typically degrading only from 0.0008% to 0.0017% at 1 kHz assuming that the avoidable distortions have been eliminated. Note however that the effect becomes greater as the global feedback factor is reduced. There is little deterioration at HF, where other distortions dominate.[4]

The VAS buffer is most useful when LF distortion is already low, as it removes Distortion 4, which is – or should be – only visible when grosser non-linearities have been seen to. Two equally effective ways of buffering are shown in Figures 4(e) and 4(f).

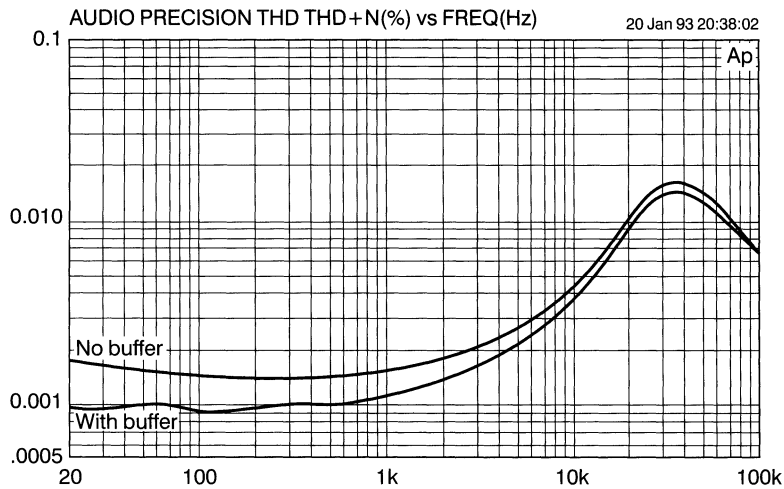



**Figure 7** *The beneficial effect of using a VAS buffer in a full scale Class B amplifier. Note that the distortion needs to be low already for the benefit to be significant.*

There are other potential benefits to VAS buffering. The effect of beta mismatches in the output stage halves is minimised.[5] Voltage drive also promises the highest $f_T$ from the output devices, and therefore potentially greater stability, though I have no data of my own to offer on this point. It is right and proper to feel trepidation about inserting another stage in an amplifier with global feedback, but since this is an emitter follower its phase shift is minimal and it works well in practice.

A VAS buffer put the right way up can implement a form of d.c. coupled bootstrapping that is electrically very similar to providing the voltage amplifier stage with a separate current-source.

The use of a buffer is essential if a VAS cascode is to do some good. Figure 7 shows before/after distortion for a full-scale power amplifier with cascode VAS driving 100 W into 8 Ω.

## Balanced voltage amplifier stage

When linearising an amplifier before adding negative feedback one of the few specific recommendations made is usually the use of a balanced voltage amplifier stage – sometimes combined with a double input stage consisting of two differential amplifiers, one complementary to the other. The latter seems to have little to recommend it, as you cannot balance a stage that is already balanced, but a balanced (and, by implication, more linear) voltage amplifier stage has its attractions. However, as explained above, the

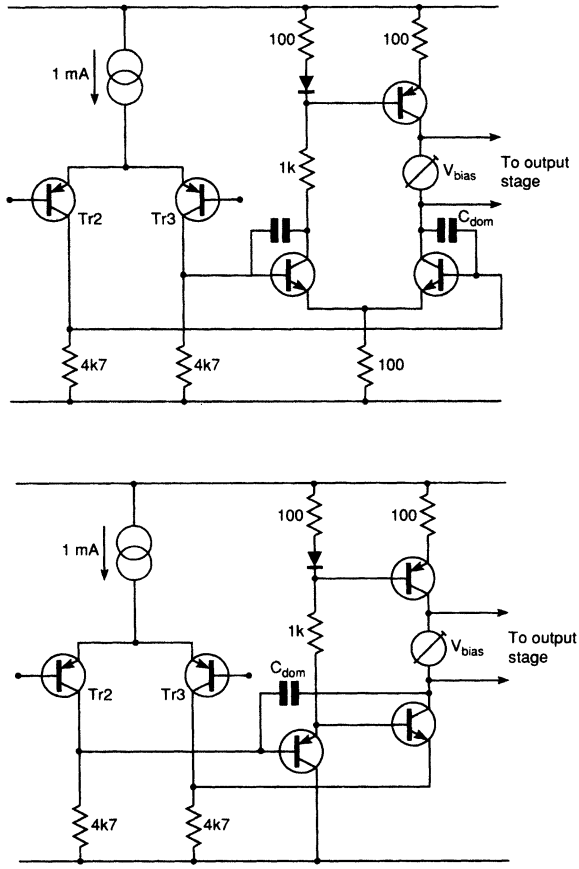



**Figure 8** *Two kinds of balanced voltage amplifier stage: Type 1 gives more open loop gain, but no better open loop linearity. Type 2 the circuit originated by Lender.*

distortion contribution from a properly designed VAS is negligible under most circumstances, and therefore there seems to be little to be gained.

Two possible versions are shown in Figure 8; The first type gives approximately 10 dB more open loop gain than the standard, but this naturally requires an increase in $C_{\mathrm{dom}}$ if the same stability margins are to be maintained. In a model amplifier, any improvement in linearity can be wholly explained by this o/l gain increase, so this seems (not unexpectedly) an unpromising approach. Also, as John Linsley Hood has pointed out,[6] the standing current through the bias generator is ill-defined compared with the usual current source VAS. Similarly the balance of the input pair is likely to be poor compared with the current-mirror version. Two signal paths from the input stage to the VAS output must have the same bandwidth; if they do not then a pole-zero doublet is generated in the open-loop

gain characteristic that will markedly increase settling-time after a transient. This seems likely to apply to all balanced voltage amplifier stage configurations. The second type is attributed by Borbely to Lender,[7] Figure 8 shows one version, with a quasi-balanced drive to the VAS transistor, via both base and emitter. This configuration does not give good balance of the input pair since it depends on the tolerances of $R_2$, $R_3$, the $V_{be}$ of the voltage amplifier stage, and so on. Borbely has advocated using two complementary versions of this giving a third type, but this would not seem to overcome the objections and the increase in complexity is significant.

All balanced voltage amplifier stages seem to be open to the objection that the vital balance of the input pair is not guaranteed, and that the current through the bias generator is not well-defined. However one advantage would seem to be the potential for sourcing and sinking large currents into $C_{dom}$, which might improve the ultimate slew-rate and HF linearity of a very fast amplifier.

## Open loop bandwidth

Acute marketing men will appreciate that reducing the LF open loop gain, leaving HF gain unchanged, must move the P1 frequency upwards, as shown in Figure 9. Open loop gain held constant up to 2 kHz appears so much better than open loop bandwidth restricted to 20 Hz. These two statements could describe near identical amplifiers, except that the first has plenty of open loop gain at LF while the second has even more. Both amplifiers have the same feedback factor at HF, where the amount available has a direct effect on distortion performance, and could easily

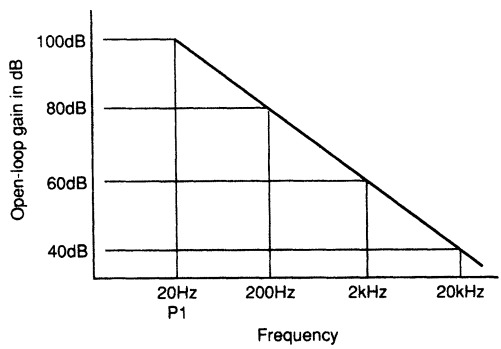



**Figure 9** *Showing how dominant pole frequency P1 can be altered by changing the LF open loop gain. The gain at HF, which determines Nyquist stability and HF distortion, is unaffected.*

have the same slew rate. Nonetheless the second amplifier somehow reads as sluggish and indolent, even when the truth of the matter is known.

Reducing low frequency open loop gain may be of interest to commercial practitioners but it also has its place in the dogma of the subjectivist. Consider it this way: firstly there is no engineering justification for it and, secondly, reducing the NFB factor will reveal more of the output stage distortion. NFB is the only weapon available to deal with this second item so blunting its edge seems ill-advised.

It is of course simple to reduce open loop gain by degenerating the input pair, but this diminishes it at HF as well as LF. To alter it at LF only requires engineering changes at the VAS. Figure 10 shows two ways. Figure 10(a) reduces gain by reducing the value of the collector impedance, having previously raised it with the use of a current source collector load. This is no way to treat a gain stage: loading resistors low enough to have a significant effect cause unwanted current variations in the VAS as well as shunting its high collector impedance, and serious LF distortion appears. While this sort of practice has been advocated in E&WW in the past,[8] it seems to have nothing to recommend it. Figure 10(b) also reduces overall open loop gain by adding a frequency insensitive component to the local shunt feedback around the voltage amplifier stage. The value of $R_{\mathrm{NFB}}$ is too high to load the collector significantly and therefore the full gain is available for local feedback at LF, even before $C_{\mathrm{dom}}$ comes into action. Figure 11 shows the effect on the open loop gain of a model amplifier for several values of $R_{\mathrm{NFB}}$; this plot is in the format described in the first part of this series where error voltage is plotted rather than gain so the curve appears upside down compared with the usual presentation. Note that the dominant pole frequency is increased from 800 Hz to above 20 kHz by using a 220 kΩ value for $R_{\mathrm{NFB}}$; however the gain at higher frequencies

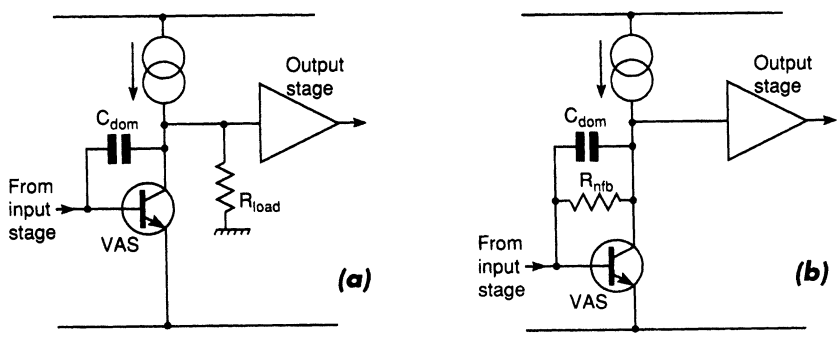



**Figure 10** *Two ways to reduce open loop gain: (a) simply loading down the collector. This is a cruel way to treat a VAS since current variations cause extra distortion. (b) local NFB with a resistor in parallel with* $C_{dom}$. *This looks crude, but actually works very well.*

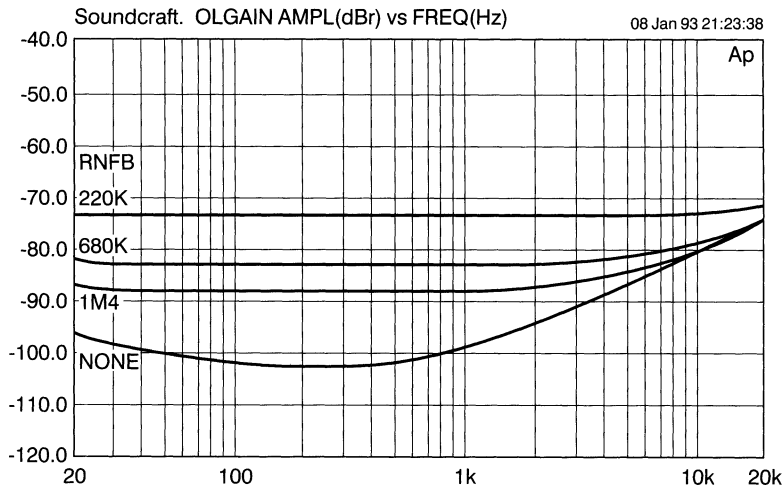



**Figure 11** *The result of voltage amplifier stage gain reduction by local feedback; the dominant pole frequency is increased from about* 800 Hz *to about* 20 kHz*, with high frequency gain hardly affected.*

is unaffected and so is the stability. Although the amount of feedback available at 1 kHz has been decreased by nearly 20 dB, the distortion at +16 dBu output is only increased from below 0.001% to 0.0013%. Most of the reading is due to noise.

In contrast, reducing the open loop gain by just 10 dB through loading the VAS collector to ground requires a load of 4.7 kΩ which, under the same conditions, yields distortion of more than 0.01%.

It might seem that the stage which provides all the voltage gain and swing in an amplifier is a prime suspect for generating the major part of its non-linearity. In actual fact, this is unlikely to be true, particularly with a cascode VAS/current source collector load buffered from the output stage. Number 2 in the distortion list can usually be forgotten.

## References

1. Gray & Meyer *Analysis & Design of Analog Integrated Circuits*, Wiley 1984, p 251 (VAS transfer characteristic).
2. Antognetti (Ed) *Power Integrated Circuits*, McGraw-Hill 1986, Section 9.3.2.
3. Gray & Meyer *Analysis & Design of Analog Integrated Circuits*, Wiley 1984, p 252 (Rco limit on VAS gain).
4. Self *Sound Mosfet Design*, *EW + WW*, September 1990, p 760.

5. Oliver *Distortion In Complementary-Pair Class-B Amplifiers, Hewlett-Packard Journal*, Feb 1971, p 11.
6. Linsley-Hood *Solid State Audio Power-3, EW + WW*, January 1990, p 16.
7. Borbely *A 60 W MOSFET Power Amplifier, The Audio Amateur*, Issue 2, 1982, p 9.
8. Hefley *High Fidelity, Low Feedback, 200 W, EW + WW*, June 92 p 45.

# 19 Distortion in power amplifiers, Part IV: the power amplifier stages

*November 1993*

The almost universal choice in semiconductor power amplifiers is for a unity gain output stage, and specifically a voltage follower. Output stages with gain are not unknown,[1] but they are not common. Most designers feel that controlling distortion while handling large currents is hard enough without trying to generate gain at the same time.

The first three parts of this series have dealt with one kind of distortion at a time, due to the monotonic transfer characteristics of small signal stages, which usually, but not invariably, work in class A.[2] Economic and thermal realities mean that most output stages are class B, and so we must now consider crossover distortion, which remains the thorniest problem in power amplifier design, and HF switchoff effects.

It is now also necessary to consider what kind of active device is to be used; jfets offer few if any advantages in the small current stages, but power fets are a real possibility, providing that the extra cost brings with it tangible benefit.

## The class war

The fundamental factor in determining output stage distortion is the class of operation. Apart from its inherent inefficiency, class A is the ideal operating mode, because there can be no crossover or switchoff distortion. However, of those designs which have been published or reviewed, it is notable that the large signal distortion produced is still significant. This looks like an opportunity lost, as of the distortion mechanisms discussed in the first part of this series, we now only have to deal with Distortion 1 (input stage), Distortion 2 (VAS), and Distortion 3 (output stage large signal non-linearity). Distortions 4, 5, 6 and 7, as mentioned earlier, are direct results

of class B operation and therefore can be thankfully disregarded in a class A design. However, class B is overwhelmingly of the greater importance, and is therefore dealt with in detail.

Class B is subject to misunderstanding. The statement is often made that a pair of output transistors operated without any bias are working in 'class B', and therefore 'generate severe crossover distortion'. In fact, with no bias each output device is operating for slightly less than half the time, and the question arises as to whether it would not be more accurate to call this class C and reserve class B for that condition of quiescent current which eliminates, or rather minimises, the crossover artifacts.

A further complication exists; it is not generally appreciated that moving into what is usually called class AB, by increasing the quiescent current, does not make things better. In fact, the THD reading will increase as the bias control is advanced, with what is usually known as '$g_m$ doubling' (i.e. a voltage gain increase caused by both devices conducting simultaneously in the centre of the output voltage range) putting edges into the distortion residual that generate high order harmonics in much the same way that underbiasing does. This important fact seems almost unknown, presumably because the $g_m$ doubling distortion is at a relatively low level and is completely obscured in most amplifiers by other distortion mechanisms.

The phenomenon is demonstrated in Figures 1(a), (b), (c) which shows spectrum analysis of the distortion residuals for under biasing, optimal, and over biasing of a 150 W/8 Ω amplifier at 1 kHz.

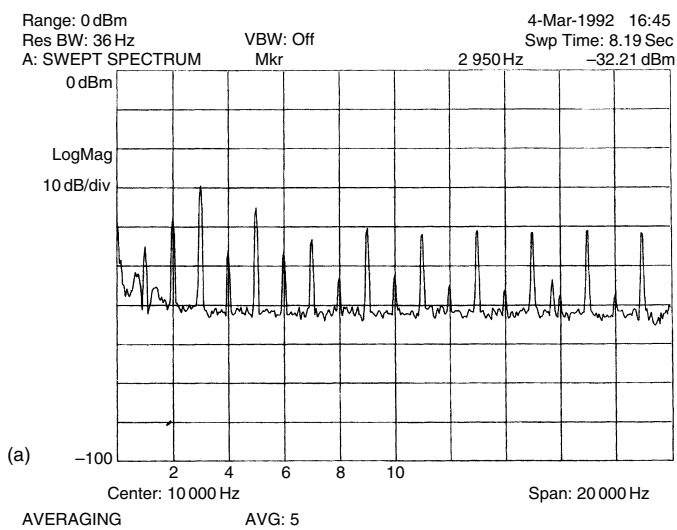
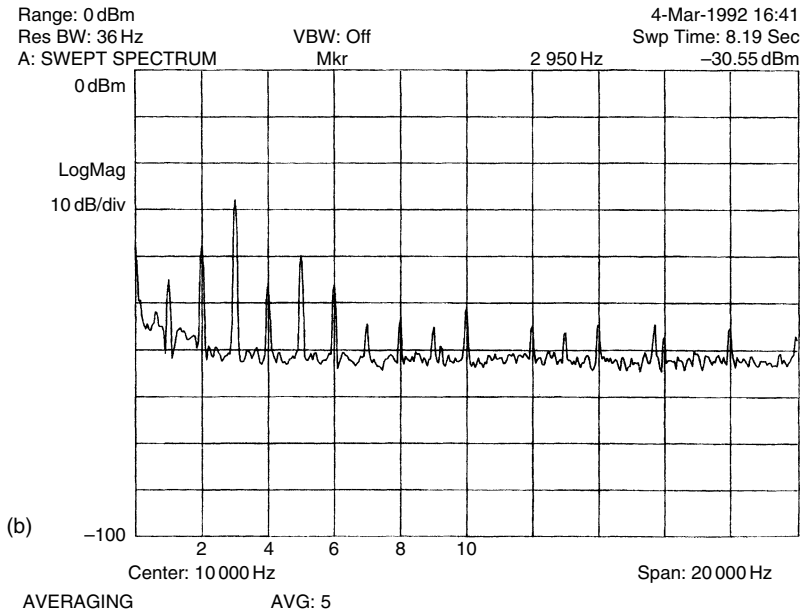
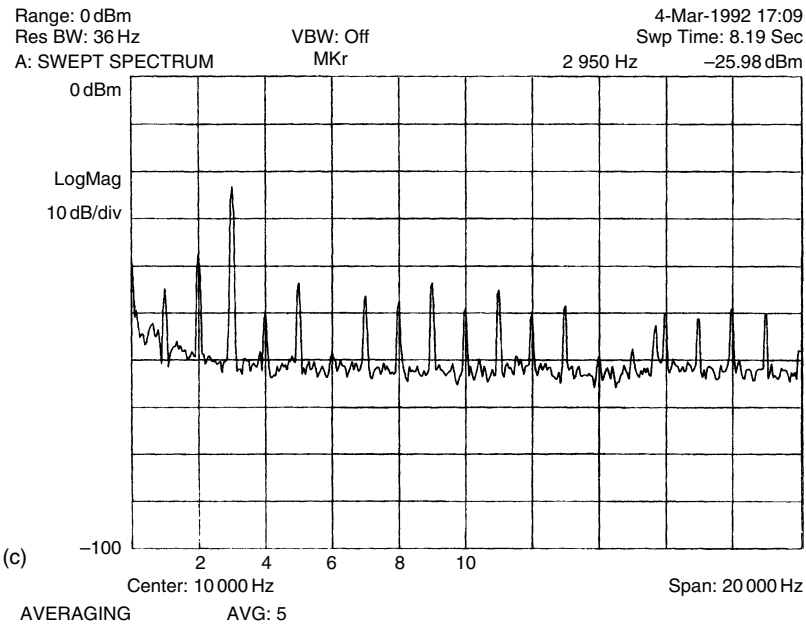



**Figure 1** *Spectrum analysis of class B & AB distortion residual. 1(a) Underblased class B; 1(b) Optimal class B; 1(c) class AB.*




(b)




(c)

**Figure 1** *Continued*

As before, all non-linearities except the unavoidable Distortion 3 (output stage) have been effectively eliminated. The over biased case had its quiescent current increased until the $g_m$ doubling edges in the residual had an approximately 50:50 mark/space ratio, and so was in class A about half the time which represents a rather generous amount of quiescent for class AB. Nonetheless, the higher order odd harmonics in Figure 1(c) are at least 10 dB greater in amplitude than those for the optimal class B case, and the third harmonic is actually higher than for the under-biased case as well. However the under biased amplifier, generating the familiar sharp spikes on the residual, has a generally greater level of high-order odd harmonics above the 5th; about 8 dB higher than the AB case.

Bearing in mind that high order odd harmonics are generally considered to be the most unpleasant, there seems to be a clear case for avoiding Class AB altogether, as it will always be less efficient and generate more high order distortion than the equivalent class B circuit, class distinction therefore seems to resolve itself into a binary choice between A or B.

It must be emphasised that these effects can only be seen in an amplifier where the other forms of distortion have been properly minimised. The r.m.s. THD reading for case **1a** was 0.00151%, for case **1b** 0.00103%, and for case **1c** 0.00153%. The tests were repeated at the 40 W power level with very similar results. The spike just below 16 kHz is interference from the test gear VDU.

This may seem complicated enough, but there are other and deeper subtleties in class B.

## Distortions of the output

I have designated the distortion produced directly by output stages as Distortion 3 (see Part 1); this subdivides into three categories. Mechanism 3a describes the large signal distortion produced by both class A and B, ultimately because of the large current swings in the active devices. In bipolars, but not fets, large collector currents reduce the beta leading to drooping gain at large output excursions. I shall use the term 'LSN' for large signal nonlinearity, as opposed to crossover and switchoff phenomena that cause trouble at all output levels.

The other two contributions to Distortion 3 are associated with class B only; Distortion 3b the classic crossover distortion resulting from the non-conjugate nature of the output characteristics, and is essentially frequency independent.

In contrast, Distortion 3c is switchoff distortion generated by the output devices failing to turn off quickly and cleanly at high frequencies. This mechanism is strongly frequency dependent. It is sometimes called switching distortion, but this allows room for confusion, as some writers use

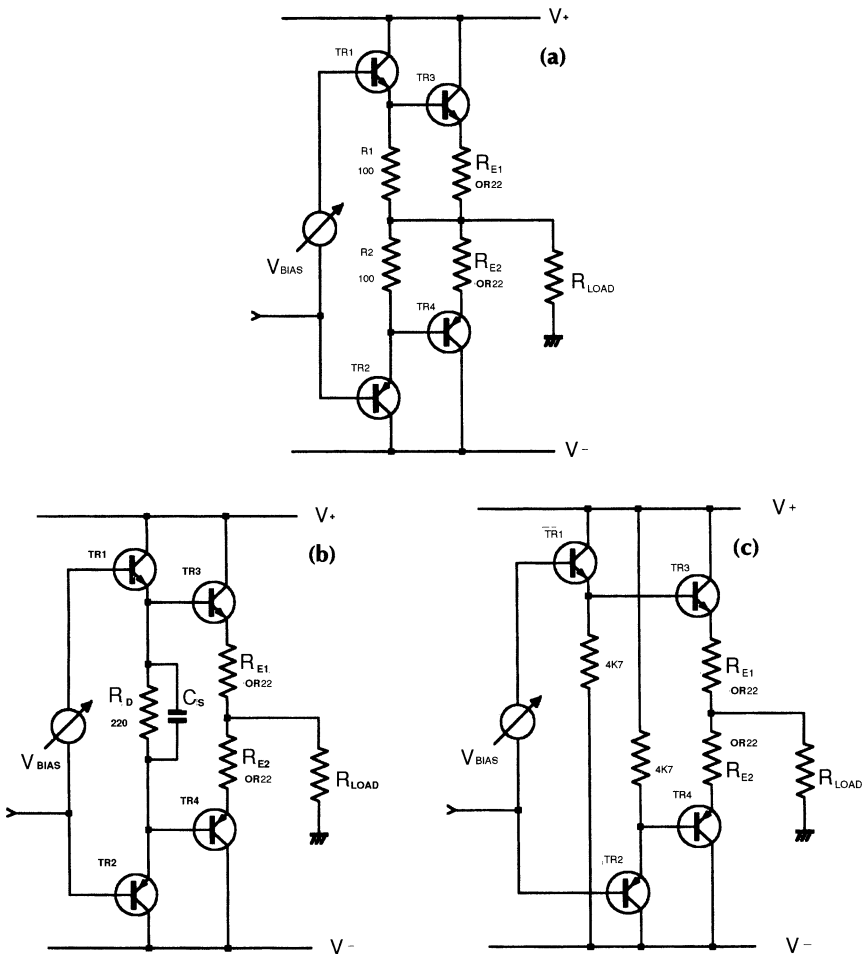



**Figure 2** *Three types of emitter follower output stages.*

'switching distortion' to cover crossover distortion as well. I refer specifically to charge storage turn off troubles.

One of my aims for this series has been to show how to isolate individual distortion mechanisms. To examine output behaviour, it is perfectly practical to drive output stages open loop providing the driving source impedance is properly specified; this is difficult with a conventional amplifier, as it means the output must be driven from a frequency dependant impedance simulating that at the VAS collector with some sort of feedback mechanism incorporated to keep the drive voltage constant.

However, if the VAS is buffered from the output stage by some form of emitter follower, as described in the last part, it makes things much

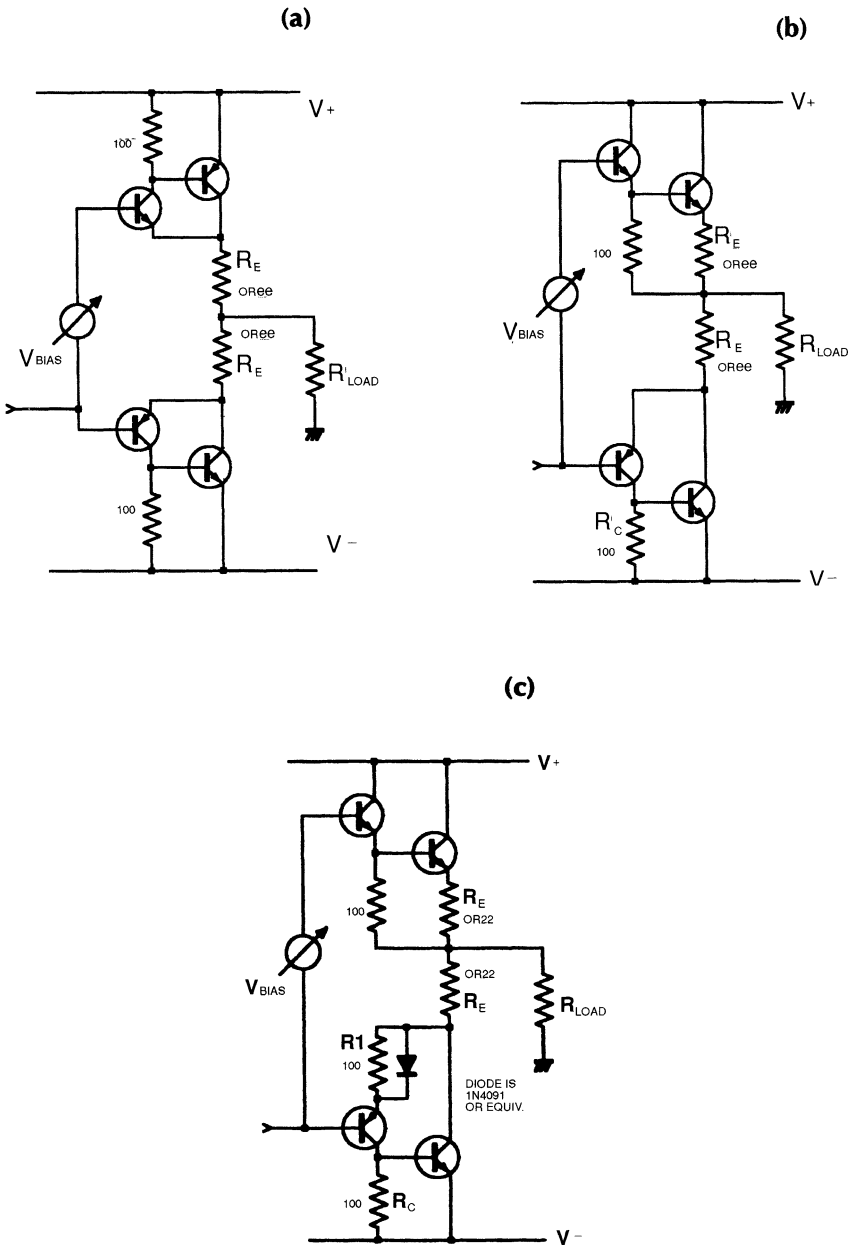



**Figure 3** *CFP circuit and quasi-complementary stages.*

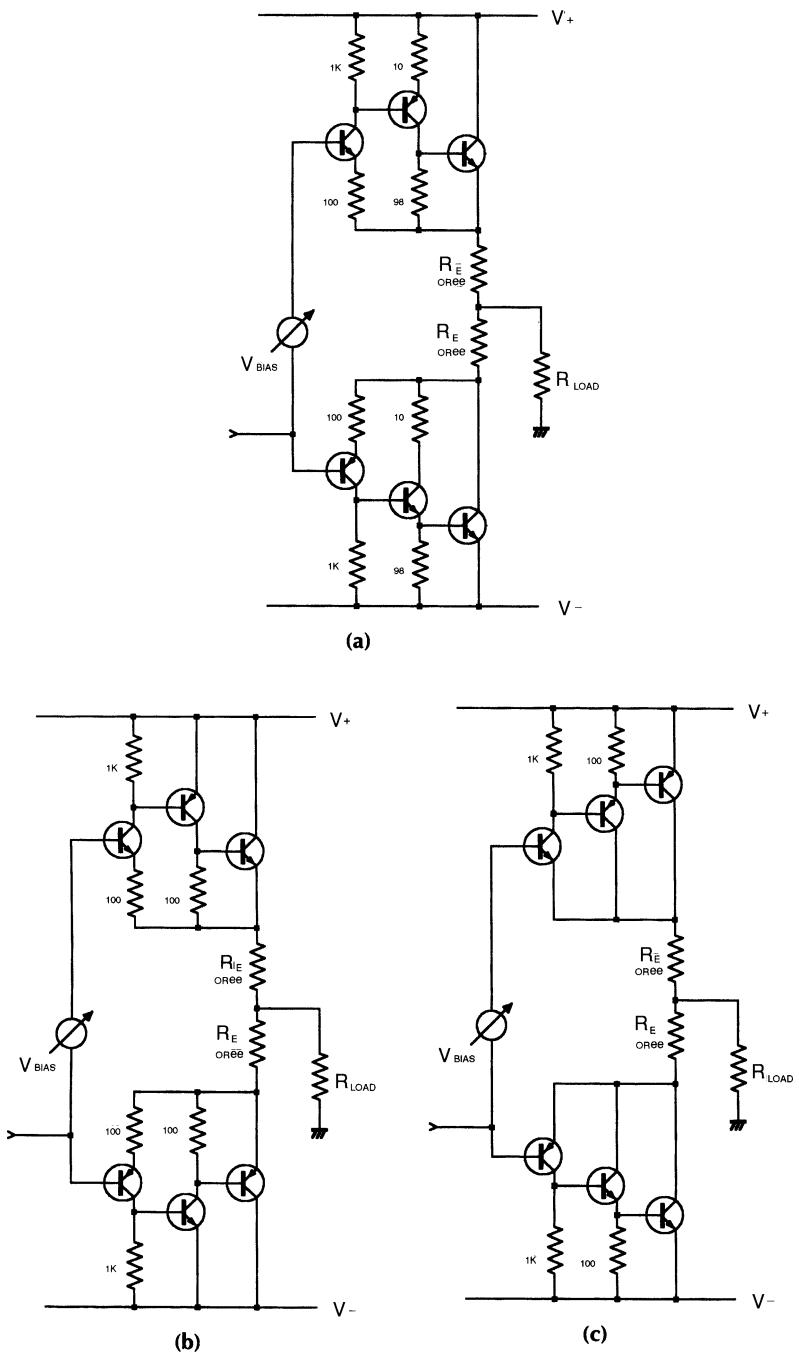



**Figure 4** *Three of the possible output triple configurations.*

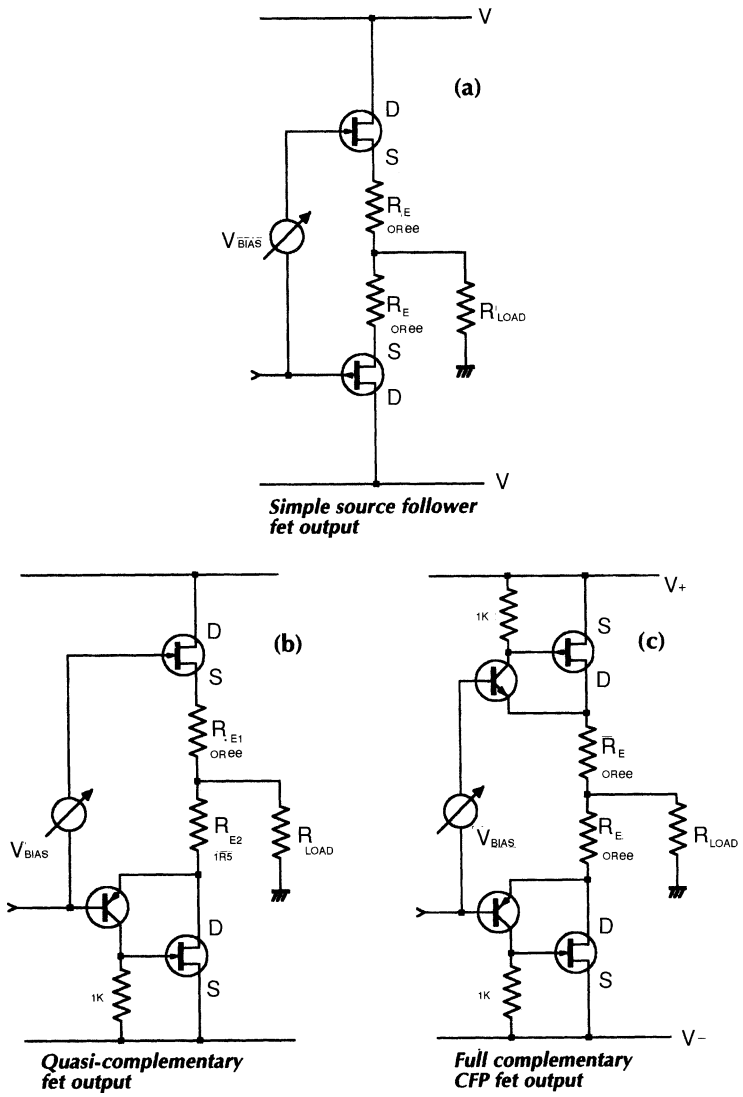



**Figure 5** *Three mosfet output architectures.*

simpler, a straightforward low impedance source (e.g. 50 Ω) providing a good approximation of a VAS-buffered closed loop amplifier. The VAS buffer makes the system more designable by eliminating two variables – the VAS collector impedance at LF, and the frequency at which it starts to decrease due to local feedback through $C_{\mathrm{dom}}$. This markedly simplifies the study of output stage behaviour.

The large signal linearity of various kinds of open loop output stage with typical values are shown in Figures 6–15 These diagrams were all generated by spice simulation, and are plotted as incremental output gain against output voltage, with the load resistance stepped from 16 Ω to 2 Ω. The power devices are *MJ802* and *MJ4502*, which are more complementary than many transistor pairs, and minimise distracting large signal asymmetry. The quiescent current is in each case set to minimise the peak deviations of gain around the crossover point for 8 Ω loading; for the moment it is assumed that you can set this accurately and keep it where you want it. The difficulties in actually doing this will be examined later.

There are at least 16 distinct configurations in straightforward output stages not including error correcting,[3] current dumping[4] or Blomley[5] types. These are as follows:

| | | |
|---|---|---|
| Emitter Follower | 3 types | Figure 2 |
| Complementary Feedback Pair | 1 type | Figure 3 |
| Quasicomplementary | 2 types | Figure 3 |
| Output Triples | At least 7 types | Figure 4 |
| Power FET | 3 types | Figure 5 |

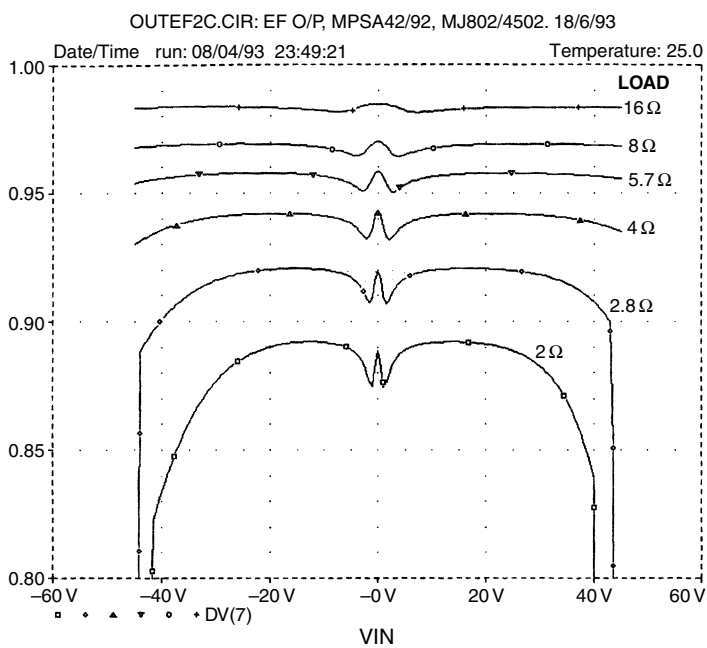



**Figure 6** *Emitter follower large signal gain vs output.*

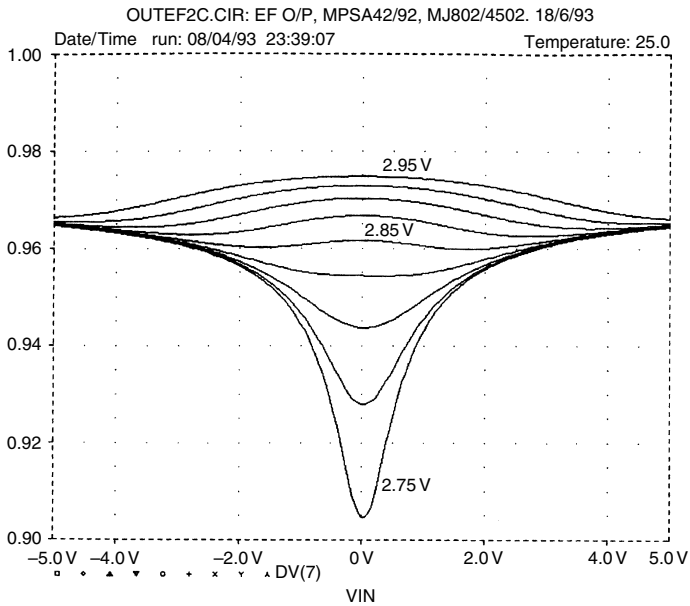



**Figure 7** *Emitter follower crossover region gain deviations, ±5 V range.*

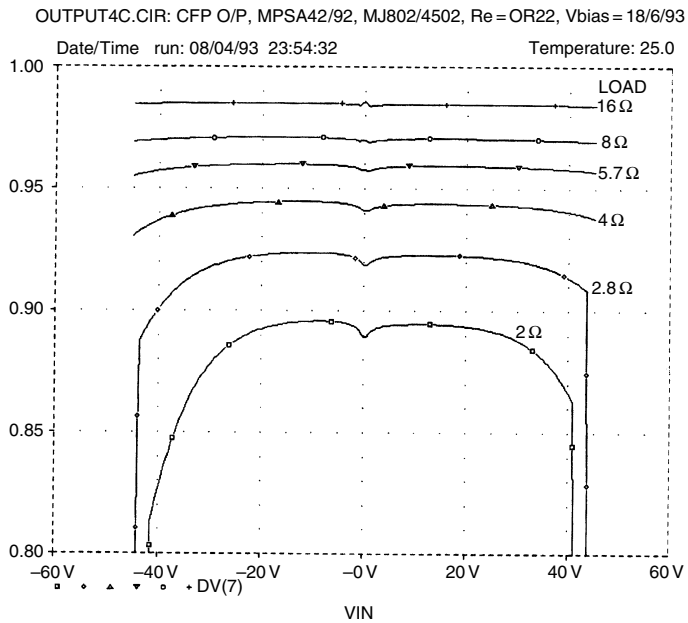



**Figure 8** *Complementary feedback pair gain versus output.*

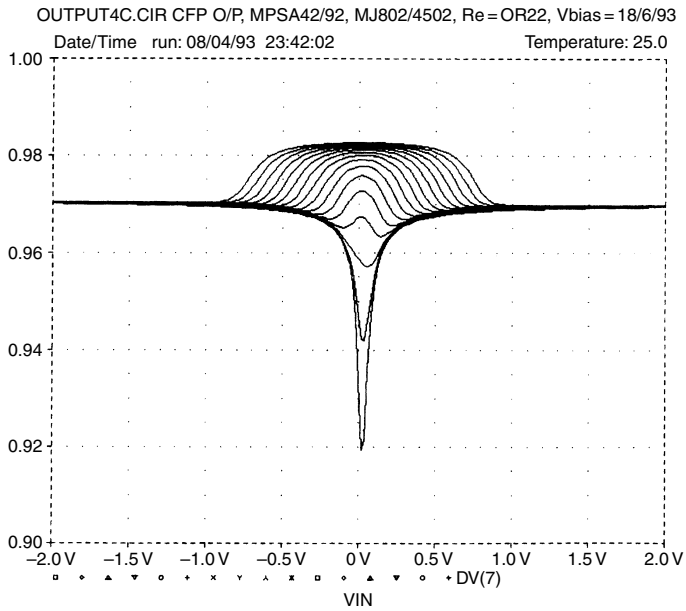



**Figure 9** *CFP crossover region ±2 V, $V_{bias}$ as a parameter.*

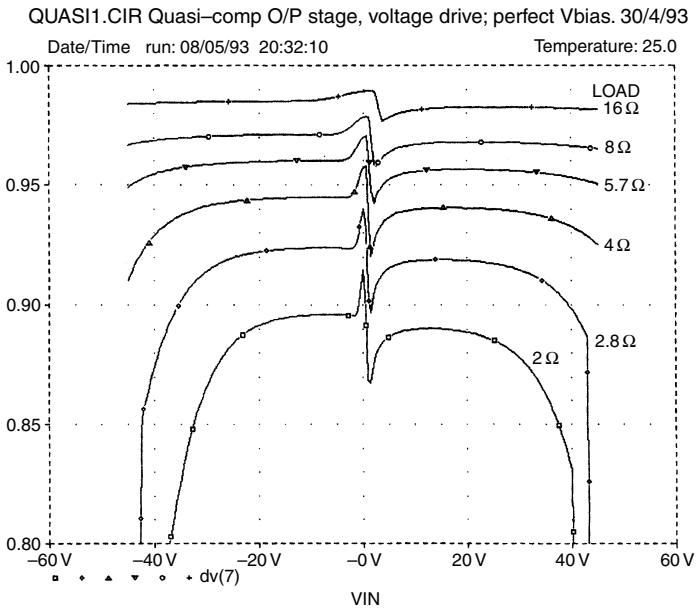



**Figure 10** *Quasi complementary large signal gain vs output load resistance.*

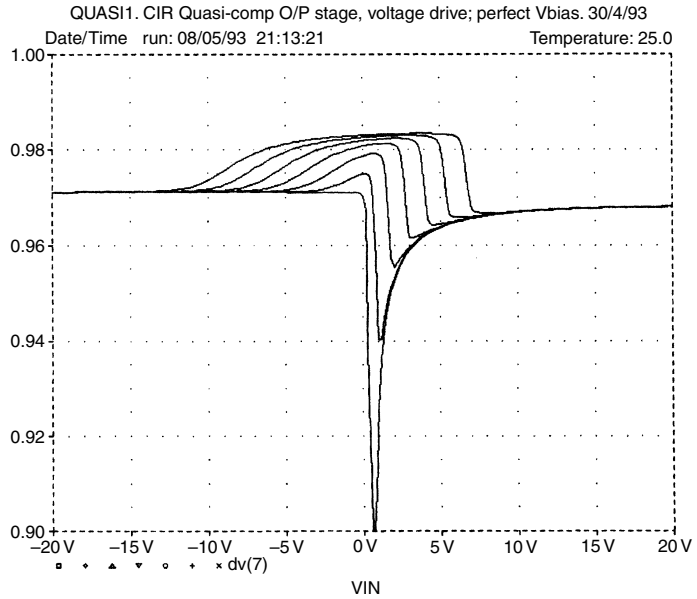



**Figure 11** *Quasi crossover region* ±20 V, $V_{bias}$ *as parameter.*

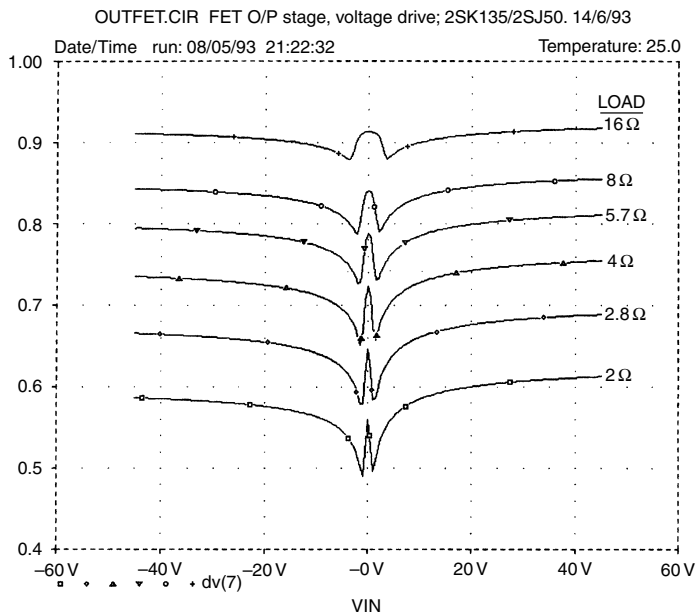



**Figure 12** *Source follower FET large signal gain vs output.*

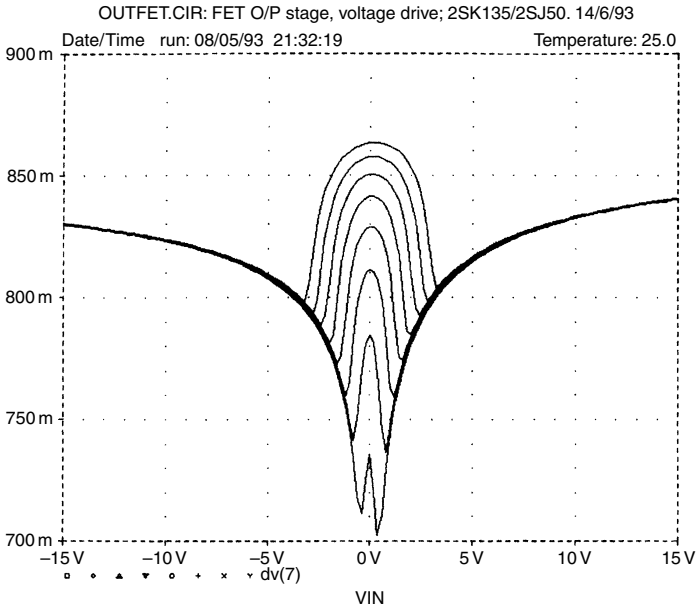



**Figure 13** *Source follower FET crossover region* ±15 V *range.*

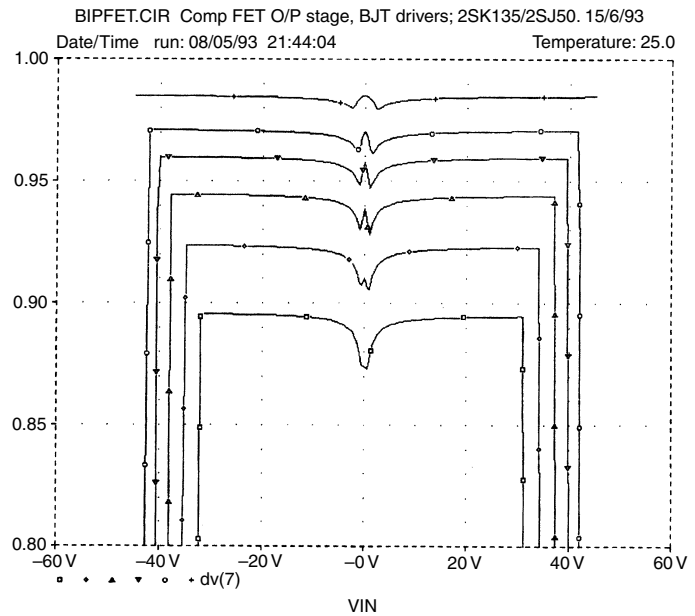



**Figure 14** *Complementary bipolar FET gain vs output.*

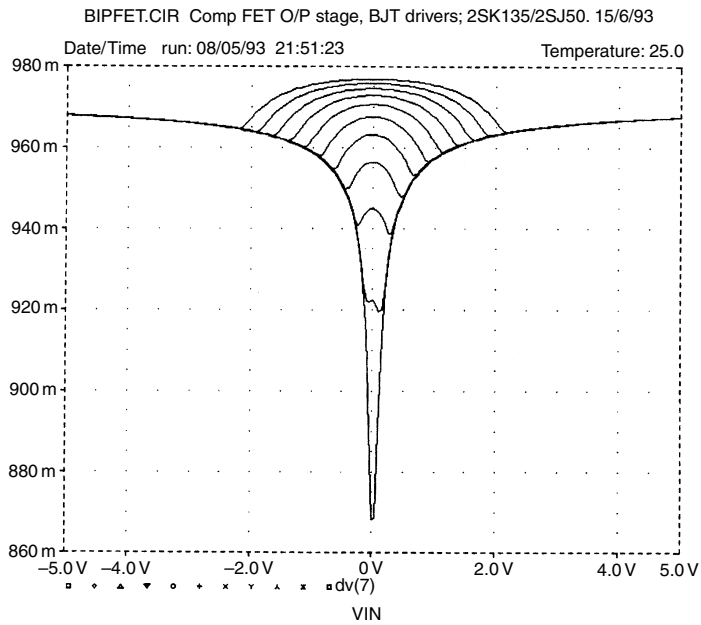



**Figure 15** *Complementary BJT FET crossover region* ±15 V *range.*

## The emitter follower output

Figure 2 shows three versions of the most common type of output stage; the double-emitter follower where the first follower acts as driver to the second (output) device. I have deliberately called this an emitter follower rather than a Darlington configuration, as this latter implies an integrated device with associated resistors. As for all the circuitry here, the component values are representative of real practice.

Two important attributes of this topology are:

- The input is transferred to the output via two base emitter junctions in series, with no local feedback around the stage (apart from the very local 100% voltage feedback that makes an emitter follower what it is);
- There are two dissimilar base emitter junctions between the bias voltage and the emitter resistor $R_e$, carrying different currents and at different temperatures. The bias generator must attempt to compensate for both at once, though it can only be thermally coupled to one. The output devices have substantial thermal inertia and thus thermal compensation represents a time average of the preceding conditions. Figure 2(a) shows the most prevalent version (type I) which has its driver emitter resistors connected to the output rail.

The type II configuration in Figure 2(b) is at first sight merely a pointless variation on type I, but in fact it has a valuable extra property. The shared driver emitter resistor $R_d$, with no output rail connection, allows the drivers to reverse bias the base emitter junction of the output device being turned off.

Assume that the output voltage is heading downwards through the crossover region; the current through $R_{e1}$ has dropped to zero, but that through $R_{e2}$ is increasing, giving a voltage drop across it, so $Tr_4$ base is caused to go more negative to get the output to the right voltage. This negative excursion is coupled to $Tr_3$ base through $R_d$, and with the values shown can reverse bias it by up to 0.5 V, increasing to 1.6 V with a 4 Ω load. The speed up capacitor $C_s$ markedly improves this action, preventing the charge suckout rate being limited by the resistance of $R_d$. While the type I circuit has a similar voltage drop across $R_{e2}$, the connection of the mid point of $R_1$, $R_2$ to the output rail prevents this from reaching $Tr_3$ base; instead $Tr_1$ base is reverse biased as the output moves negative, and since charge storage in the drivers is usually not a problem, this does little good. In the type II circuit the drivers are never reverse biased, though they do turn off. The important issue of output turn off and switching distortion is further examined in the next part of this series.

The type III topology shown in Figure 2(c) maintains the drivers in class A by connecting the driver emitter resistors to the opposite supply rail rather than the output rail. It is a common misconception[6] that class A drivers somehow maintain better low frequency control over the output devices, but I have yet to substantiate any advantage myself. The driver dissipation is of course substantially increased, and nothing seems to be gained at LF as far as the output transistors are concerned, for in both type I and type II the drivers are still conducting at the moment the outputs turn off, and are back in conduction before the outputs turn on, which would seem to be all that matters.

Type III is equally good as type II in reverse biasing the output bases, and may give even cleaner HF turn off as the carriers are being swept from the bases by a higher resistance terminated in a higher voltage approximating constant current drive; I have yet to try this.

The large signal linearity of the three versions is virtually identical – all have the same feature of two base emitter junctions in series between input and load.

The gain/output voltage plot is shown at Figure 6; with BJTs the gain reduction with increasing loading is largely due to the emitter resistors. Note that the crossover region appears as a relatively smooth wobble rather than a jagged shape. Another major feature is the gain droop at high output voltages and low loads indicating that high collector currents are the fundamental cause of this.

A close up of the crossover region gain for 8 Ω loading only is shown in Figure 7; note that no $V_{bias}$ setting can be found to give a constant or even monotonic gain; the double dip and central gain peak are characteristic of optimal adjustment. The region extends over about ±5 V, independent of load resistance.

## Complementary feedback output

The other major type of bipolar output is the complementary feedback pair (CFP) sometimes called the Sziklai Pair, Figure 3(a). There seems to be only one popular configuration, though versions with gain are possible. The drivers are now placed so that they compare output voltage with that at the input. Wrapping the outputs in a local negative feedback loop promises better linearity than emitter follower versions with 100% feedback applied separately to driver and output transistors.

This topology also has better thermal stability, because the $V_{be}$ of the output devices is inside the local feedback loop, and only the driver $V_{be}$ affects the quiescent current. It is usually simple to keep drivers cool, and thermal feedback from them to the $V_{bias}$ generator transistor can be much faster and mechanically simpler.

Like emitter follower outputs, the drivers are conducting whenever the outputs are, and so special arrangements to keep them in class A seem pointless. This stage, like emitter follower type I, can only reverse bias the driver bases rather than the outputs, unless extra voltage rails outside the main ones are provided.

The output gain plot is shown in Figure 8. Fourier analysis shows that the CFP generates less than half the large signal distortion of an emitter follower stage. (See Table 1) Given also the greater quiescent stability, it is hard to see why this topology is not more popular.

The crossover region is much narrower, at about ±0.3 V (Figure 9). When under biased, this shows up on the distortion residual as narrower spikes than an emitter follower output gives.

The bad effects of $g_m$ doubling as $V_{bias}$ increases above optimal (here 1.296 V) can be seen in the slopes moving outwards from the centre.

## Quasicomplementary outputs

The original quasicomplementary configuration[7] was almost mandatory, as it was a long time before pnp silicon power transistors matched the performance of the npn versions. The standard version shown at Figure 3(b) is well known for poor symmetry around the crossover region, as shown at Figure 10. A close-up of the crossover region (Figure 11) reveals an

unhappy hybrid of the emitter follower and CFP, as might be expected, and that no setting of bias voltage can remove the sharp edge in the gain plot.

A major improvement to symmetry may be made by using a Baxandall diode[8] as shown in Figure 3(c). This stratagem yields gain plots very similar to those for the true complementary emitter follower at Figures 6, 7, though in practice the crossover distortion seems rather higher. When a quasiBaxandall stage is used closed loop in an amplifier in which distortion mechanisms 1 and 2, and 4 to 7 have been properly eliminated, it is capable of better performance than is commonly believed. For example, 0.0015% (1 kHz) and 0.015% (10 kHz) at 100 W is straightforward to obtain from an amplifier with a negative feedback factor of about 34 dB at 20 kHz.

The best reason to use the quasiBaxandall approach today is to save money on output devices, as pnp power transistors remain somewhat pricier than npns. Given the tiny cost of a Baxandall diode, and the absolutely dependable improvement it gives, there seems no reason why anyone should ever use the standard quasi circuit. My experiments show that the value of $R_1$ in Figure 3(c) is not critical; making it about the same as $R_c$ seems to work well.

## Triples

With three rather than two bipolar transistors in each half of an output stage the number of circuit permutations possible leaps upwards. There are two main advantages if output triples are used correctly: better linearity at high output voltages and currents; and more stable quiescent setting as the predrivers can be arranged to handle very little power, and remain almost cold in use.

However, triples do not automatically reduce crossover distortion, and they are, as usually configured, incapable of reverse biasing the output bases to improve switch-off. Figure 4 shows three ways to make a triple output stage – all of those shown (with the possible exception of Figure 4(c), which I have just made up) have been used in commercial designs. The circuit of 4a is the Quad *303* quasicomplementary triple. The design of triples demands care, as the possibility of local HF instability in each output half is very real.

## Power FET outputs

Power mosfets are often claimed to be a solution to all amplifier problems, but they have their drawbacks: poor linearity and a high on-resistance that makes output efficiency mediocre. The high frequency response is better, implying that the second pole P2 of the amplifier response will be higher,

allowing the dominant pole P1 be raised with the same stability margin, and in turn allowing more overall feedback to reduce distortion. However, the extra feedback (if it proves available in practice) is needed to correct the higher open loop distortion.

To complicate matters, the compensation cannot necessarily be lighter because the higher output resistance makes the lowering of the output pole by capacitive loading more likely.

The extended frequency response creates its own problems; the HF capabilities mean that rigorous care must be taken to prevent parasitic oscillation, as this is often promptly followed by an explosion of disconcerting violence. Fets should at least give freedom from switchoff troubles as they do not suffer from charge storage effects.

Three types of FET output stage are shown in Figure 5. Figures 12 to 15 show spice gain plots, using *2SK135/2SJ50* devices.

Most FET amplifiers use the simple source follower configuration in Figure 5(a); the large signal gain plot at Figure 12 shows that the gain for a given load is lower, (0.83 rather than 0.97 for bipolar, at 8 Ω) because of low $g_m$. This, with the high on resistance, noticeably reduces output efficiency.

Open loop distortion is markedly higher; however large signal nonlinearity does not increase with heavier loading, there being no equivalent of 'bipolar gain droop'. The crossover region has sharper and larger gain deviations than a bipolar stage, and generally looks pretty nasty; Figure 13 shows the difficulty of finding a 'correct' $V_{\text{bias}}$ setting.

Figure 5(b) shows a hybrid (i.e. bipolar/FET) quasi complementary output stage.[9] The stage is intended to maximise economy rather than performance, once the decision has been made (probably for marketing reasons) to use fets, by making both output devices cheap n-channel devices; complementary mosfet pairs remain relatively rare and expensive.

The basic configuration is badly asymmetrical, the hybrid lower half having a higher and more constant gain than the source follower lower upper half. Increasing the value of $R_{e2}$ gives a reasonable match between the gains of the two halves, but leaves a daunting crossover discontinuity.

The hybrid full complementary stage in Figure 5(c) was conceived[10] to maximise performance by linearising the output devices with local feedback and reducing $I_q$ variations due to the low power dissipation of the bipolar drivers. It is highly linear, showing no gain droop at heavier loadings (Figure 14) and promises freedom from switchoff distortion. But, as shown, it is rather inefficient in voltage swing. The crossover region (Figure 15) still has some dubious sharp corners, but the total crossover gain deviation (0.96–0.97 at 8 Ω) is much smaller than for the quasi hybrid (0.78–0.90) and so less high order harmonic energy is generated.

## References


1. Mann, R. 'The Texan 20 + 20 watt stereo amplifier'. *Practical Wireless*, May 1972, p 48.
2. Takahashi, 'Design & construction of high slew rate amplifiers', Preprint No. 1348 (A 4) for *60th AES Convention*, 1978.
3. Hawksford MO. 'Distortion correction in audio power amplifiers', *JAES* January/February 1981 p 27 (Error correction).
4. Walker, P. 'Current dumping audio amplifier', *Wireless World*, 1975 pp 60/62.
5. Blomley. 'New approach to class B'. *Wireless World*, February 1971, p 57 and March 1971, p 127/131.
6. Otala. 'An audio power amplifier for ultimate quality requirements', *IEEE Trans on Audio & Electroacoustics*, December 1973, p 548.
7. Lin, H. *Electronics*, September 1956 pp 173/175.
8. Baxandall, P. 'Symmetry in class B. Letters', *Wireless World*, September 1969 p 416.
9. Self, D. 'Sound mosfet design'. *Electronics & Wireless World*, September 1990, p 760.
10. Self D. 'Mosfet audio output', letter, *Electronics & Wireless World*, May 1989, p 524 (see also Ref. 9).

# 20 Distortion in power amplifiers, Part V: output stages

*December 1993*

From earlier work in this series, distortion from the small-signal stages may be kept to levels that will prove negligible compared with distortion from a closed-loop output stage. Similarly, future work in this series will show that distortion mechanisms 4 to 7 from my original list (*EW + WW*, *July 93*) can be effectively eliminated by lesser-known but straightforward methods. This leaves the third mechanism in its three components as the only distortion that is in any sense unavoidable: Class-B stages free from crossover artifacts are not exactly commonplace.

This is a good place to introduce the concept of a blameless amplifier, one designed so that all the easily-defeated distortion mechanisms have been rendered negligible. The word *blameless* has been carefully chosen to not imply perfection.

The first distortion, non-linearity in the input stage, cannot be totally eradicated but its onset can be pushed well above 20 kHz. The second distortion, non-linearity in the voltage amplifier stage, can be effectively eliminated by cascoding. Distortion mechanisms 4 to 7, concerned with such things as earth return loops, power supply impedance and non-linear loading, can be made negligible by simple measures to be described later.

## Large-signal distortion

The large-signal nonlinearity performance of all the bipolar junction transistor stages outlined in the previous part of this series have these features in common:

Large-signal nonlinearity increases as load impedance decreases. In a typical output stage loaded with 8 Ω, closed-loop LSN is usually

negligible, the THD residual being dominated by high-order crossover artefacts that are reduced less by negative feedback. At lower impedances, such as 4 Ω, relatively pure third harmonic becomes obvious in the residual.

LSN worsens as the driver emitter or collector resistances are reduced, because the driver current swings are larger. On the other hand, this reduction improves output device turn-off, and will so decrease switchoff distortion; the usual compromise is around 47–100 Ω.

The BJT output gain plots in the previous article reveal that the LSN is compressive, the voltage gain falling off with higher output currents. It is roughly symmetrical, generating third-harmonic, and is much greater at the very lowest load impedances; this is more of an issue now that 2 Ω-capable (for a few minutes, anyway) amplifiers are considered macho, and some speaker designers are happy with 2 Ω impedance troughs.

I suggest that the fundamental reason for this gain droop is the fall in output transistor beta as collector current increases, due to the onset of high level injection effects.[1] In the emitter follower topology, this fall in beta draws more output transistor base current from the driver emitter, pulling its gain down further from unity; this is the change in gain that affects the overall transfer ratio.

The output device gain is not directly affected, as beta does not appear in the classical expression for emitter follower gain, providing the source impedance is negligibly low. This assertion has been verified by altering an output stage simulated in Spice such that the output bases are driven directly from zero impedance voltage sources rather than drivers; this abolishes the gain droop effect, so it must be in the drivers rather than the output transistors.

Further evidence for this view is that in Spice simulation, the output device Ebers-Moll model can be altered so that beta does not drop with $I_c$ (simply increase the value of the parameter IKF) and once more the gain droop does not occur, even with drivers. Here is one of the best uses of circuit simulation tweaking the untweakable. Gain droop does not affect FET outputs, which have no equivalent beta loss mechanism. See Figure 12 of chapter 20, where the wings of the FET gain plot do not turn downwards at large outputs.

It used to be commonplace for output transistors to be sold in pairs roughly matched for beta, allegedly to minimise distortion; this practice seems to have been abandoned. Simulation shows that beta mismatch produces an unbalanced gain droop that markedly increases low order harmonics without much effect on the higher ones. Modern amplifiers with adequate feedback factors will linearise this effectively. This appears to be why the practice has ceased.

## Improving large signal linearity

It will be suggested that, in a closed loop blameless amplifier, the large signal nonlinearity contribution to total distortion (for 8 Ω loading) is actually very small compared with that from crossover and switchoff. This is no longer true at 4 Ω and still less so for lower load impedances. Thus ways of reducing this mechanism will still be useful.

The best precaution is to choose the most linear output topology; The previous article suggested that the open loop complementary feedback pair output is at least twice as linear as its nearest competitor, (the emitter follower output) and so the CFP is usually the best choice unless the design emphasis is on minimising switchoff distortion.

In the small signal stages, we could virtually eliminate distortion. If the linearity of the input or voltage amplifier stage was inadequate, it was possible to come up with several ways in which it could be dramatically improved. A Class B output stage is a tougher proposition. In particular we must avoid complications to the forward path that lower the second amplifier pole P2, as this would reduce the amount of feedback that can be safely applied.

Several authors[2,3] have tried to show that the output emitter resistors of bipolar outputs can be fine tuned in value to minimise large signal distortion, the rationale being that the current dependent internal $r_e$ of the output transistors will tend to cause the gain to rise at high currents, and that this gain variation can be minimised by appropriate choice of the external $R_e$. This is not true in practical output stages whose gain behaviour tends to be dominated by beta loss and its effect on the drivers. In any case the resistor values suggested are such tiny fractions of an ohm that quiescent stability would be perilous.

In real life the $R_e$ of a CFP output stage can be varied between 0.5 and 0.2 Ω without significantly affecting linearity; 0.22 Ω is a good compromise between efficiency and stability.

The gain droop at high $I_c$s can be partly cancelled by a simple but effective feedforward mechanism. The emitter resistors $R_e$ are shunted with silicon power diodes, which with typical, circuit values will only conduct when 4 Ω loads (or less) are driven. This causes a slight gain increase that works against the beta loss droop. The modest but dependable improvement can be seen in Figure 1, measured with a 2.7 Ω load.

If a 100 W/8 Ω amplifier is required to drive 4 Ω loads then it will need paralleled output devices to cope with the power dissipation. Perhaps surprisingly, the paralleling of output BJTs (driven as usual from a single driver) has little effect on linearity, given elementary precautions to ensure current sharing. However, for the 2 Ω case there is a definite linearity improvement on resorting to tripled output devices; this is consistent with the theory that LSN results from beta loss at high collector currents.

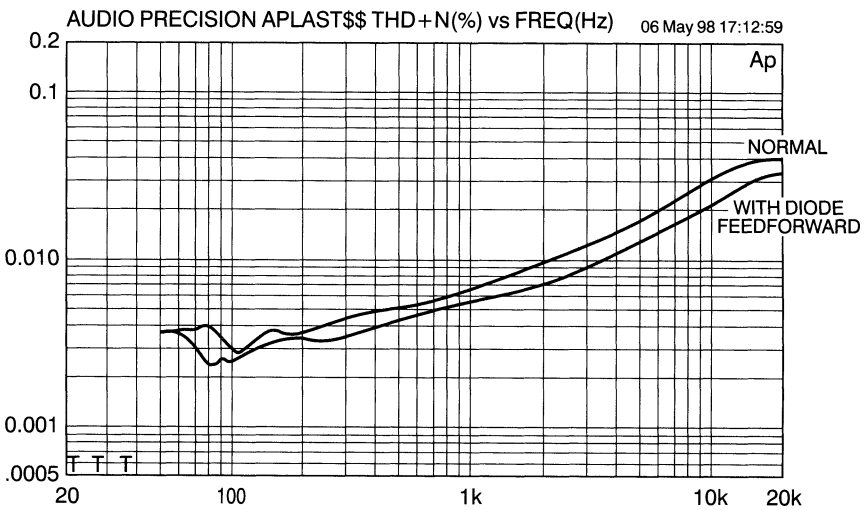



**Figure 1** *Simple diode feedforward reduces distortion with sub-*8 Ω *loads. Measured at* 210 W *into* 2.7 Ω.

## Crossover distortion

The worst problem in Class B is the crossover region, where control of the output voltage must be transferred from one device to another. Crossover distortion generates unpleasant high order harmonics with the potential to increase in percentage as signal level falls. There is a consensus that crossover caused the transistor sound of the 1960's, though to the best of my knowledge this has never actually been confirmed by the double blind testing of vintage equipment.

The $V_{bc}$-$I_c$ characteristic of a bipolar transistor is initially exponential, blending into linear as the emitter resistance $R_e$ comes to dominate the transconductance. The usual Class B stage puts two of these curves back to back, and Peter Blomley has shown that these curves are non-conjugate,[4] i.e. there is no way they can be rearranged to sum to a completely linear transfer characteristic, whatever the offset imposed by the bias voltage.

This can be demonstrated quickly and easily by Spice simulation; see Figure 2. There is at first sight not much you can do except maintain the bias voltage, and hence quiescent current, at some optimal level for minimum gain deviation at crossover; quiescent current control is a topic that could fill a book in itself, and cannot be considered properly here.

It should be said that the crossover distortion levels generated in a blameless amplifier can be low up to 1 kHz, being barely visible in residual noise and only measurable with a spectrum analyser. For example, if a blameless closed-loop Class B amplifier is driven through a *TL072* unity

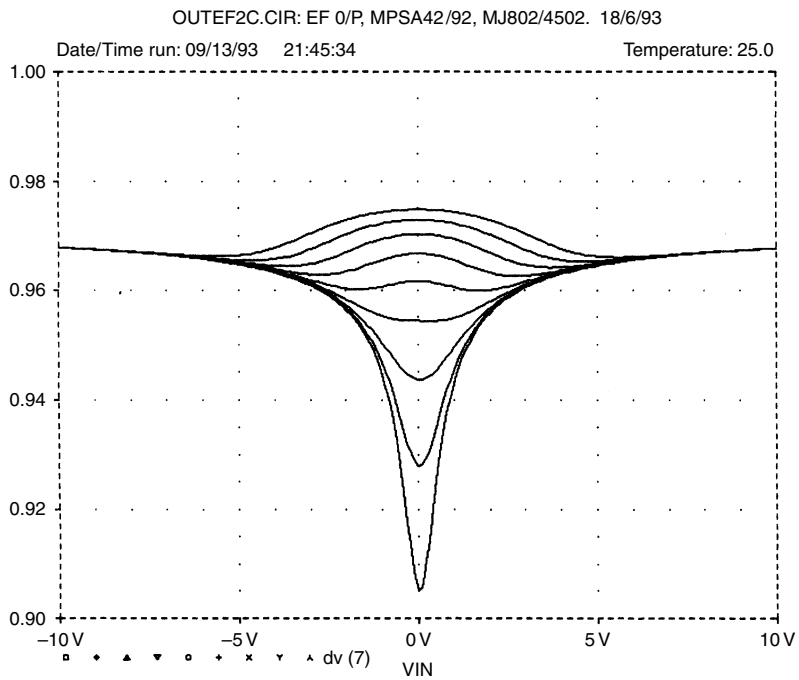



**Figure 2** *Gain/output-voltage Spice plot for an emitter follower output shows how non-conjugate transistor characteristics at the crossover region cannot be blended into a flat line at any bias voltage setting. Bias varies* 2.75 *to* 2.95 V *in* 25 mV *steps, from too little to too much quiescent.*

gain buffer the added noise from this op-amp will usually submerge the 1 kHz crossover artifacts into the noise floor. (It is most important to note that distortion mechanisms 4 to 7 create disturbances of the THD residual at the zero crossing point that can be easily mistaken for crossover distortion, but the actual mechanisms are quite different.) However, the crossover distortion becomes obvious as the frequency increases, and the high order harmonics benefit less from NFB. See text panel *Harmonic generation by crossover distortion* (p. 197).

It will be seen later that in a blameless amplifier the linearity is dominated by crossover distortion, even with a well designed and optimally biased output stage. There is an obvious incentive to minimise it, but there seems no obvious way to reduce crossover gain deviations by tinkering with any of the relatively conventional stages considered so far. Significant improvement is only likely through application of one of the following techniques:

- The use of Class AB stages where the handover from one output device to the other is genuinely gradual, and not subject to the $g_m$ doubling

effects that an over biased Class B stage shows. One possibility is the so called Harmonic AB mode.[5]

- Non-switching output stages where the output devices are clamped to prevent turn off, and thus hopefully avoiding the worst part of the $V_{bc}$-$I_c$ curve.[6]
- Error correcting output stages implementing either error feedforward or error feedback. The latter is not the same thing as global NFB, being instead a form of cancellation.[7]

Once more, these will have to be examined in the future.

## Switching distortion

This depends on several variables, notably the speed characteristics of the output devices and the output topology. Leaving aside the semiconductor physics and concentrating on the topology, the critical factor is whether or not the output stage can reverse bias the output device base emitter junctions to maximise the speed at which carriers are sucked out, so the device is turned off quickly.

The only conventional configuration that can reverse bias the output base emitter junctions is the emitter follower *type II*, described in the previous article. A second influence is the value of the driver emitter or collector resistors; the lower they are the faster the stored charge can be removed.

Applying these criteria can reduce HF distortion markedly, but it is equally important that it minimises output conduction overlap at high frequencies. If unchecked, overlap results in an inefficient and potentially destructive increase in supply current.[8] Illustrating this, Figure 3 shows current consumption vs frequency for varying driver collector resistance, for a CFP type output.

Figure 4 shows how HF THD is reduced by adding a speed-up capacitor over the common driver resistor of a *type II* emitter follower. At LF the difference is small, but at 40 kHz THD is halved, indicating much cleaner switch-off. There is also a small benefit over the range 300 Hz to 8 kHz.

## Selecting an output stage

Even if we stick to the most conventional of output stages, there are still an embarrassingly large number to choose from. The cost of a complementary pair of power fets is currently at least twice that of roughly equivalent BJTs, and taken with the poor linearity and low efficiency of these devices, the use of them may require a marketing rather than a technical motivation.

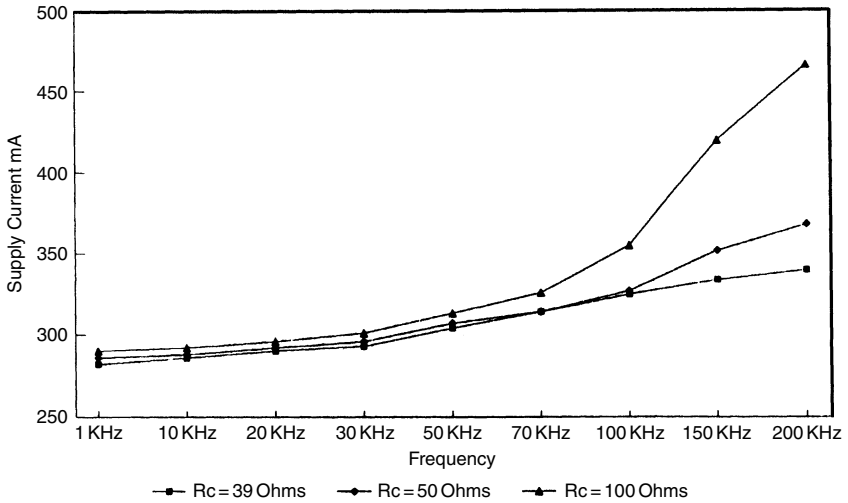



**Figure 3** *Power supply current versus frequency, for a CFP output with the driver collector resistors varied. There is little to be gained from reducing* $R_c$ *below* 50 Ω.

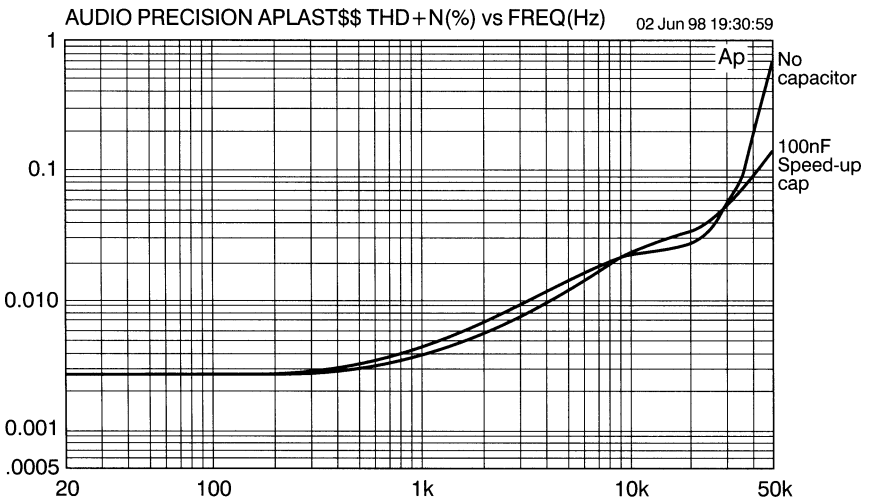



**Figure 4** *HF THD reduction by adding speed-up capacitance across the common driver resistance of a Type 11 emitter follower output stage. Taken at* 30 W/8 Ω.

Turning to BJTs, and taking the material in this article with that in *Part 4*, I conclude that these are the following candidates for best output stage:

- The emitter follower *type II* output stage is the best at coping with switchoff distortion but the quiescent current stability is not of the best;

- The CFP topology has good quiescent stability and low LSN; its worst drawback is that reverse biasing the output bases for fast switchoff is impossible without additional HT rails;
- The quasi-complementary with Baxandall diode stage comes close to mimicking the emitter follower type stages in linearity, with a potential for cost saving on output devices. Quiescent stability is not as good as the CFP.

## Closing the loop

In chapters 17 and 18, it was shown how relatively simple design rules could ensure that the THD of the small signal stages alone could be reduced to less than 0.001% across the audio band, in a repeatable fashion, and without using frightening amounts of negative feedback. Combining this subsystem with one of the more linear output stages such as the CFP version which gives 0.014% THD open loop, and having a feedback factor of at least 70 times across the band, it seems we have the ingredients for a virtually distortionless power amplifier, with THD below 0.001% from 10 Hz to 20 kHz. However, life is rarely so simple. . . .

### The seven main sources of distortion

It is one of the central themes of this series that the primary sources of power amplifier distortion are seven-fold:

1 Nonlinearity in the input stage. For a well balanced differential pair distortion rises at 18 dB/octave, and is third harmonic. When unbalanced, HF distortion is higher and rises at 12 dB/octave, being mostly second harmonic.
2 Nonlinearity of the voltage amplifier stage (VAS), second harmonic, rising at 6 dB/octave.
3 Nonlinearity of the output stage. In Class B this may be a mix of large signal distortion and crossover effects, in general rising at 6 dB/octave as the amount of NFB decreases; worsens with heavier loads.
4 Nonlinear loading of the VAS by the nonlinear input impedance of the output stage. Magnitude is essentially constant with frequency.
5 Nonlinearity caused by large rail decoupling capacitors feeding the distorted supply rail signals into the signal ground.
6 Nonlinearity caused by induction of Class B supply currents into the output, ground, or negative feedback lines.
7 Nonlinearity resulting from taking the NFB feed incorrectly.

**Table 1** *Summary of closed loop amp THD performance*

| | 1 kHz (%) | 10 kHz (%) | |
|---|---|---|---|
| Emitter follower | 0.0019 | 0.013 | Figure 5 |
| CFP | 0.0008 | 0.005 | Figure 6 |
| Quasi Bax | 0.0015 | 0.015 | Figure 7 |

Figure 5 shows the distortion performance of such a closed loop amplifier with an emitter follower output stage, Figure 6 showing the same with a CFP output stage. Figure 7 shows the THD of a quasi-complementary stage with Baxandall diode.[9] In each case distortion mechanisms 1, 2 and 4–7 have been eliminated by methods described in past and future chapters, to make the amplifier blameless.

It will be seen at once that these amplifiers are definitely not distortionless, though the performance is markedly superior to the usual run of hardware. THD in the LF region is very low, well below a noise floor of 0.0007%, and the usual rise below 100 Hz is very small indeed. However, above 2 kHz, THD rises with frequency at between 6 to 12 dB/octave, and the residual in this region is clearly time aligned with the crossover region, and consists of high order harmonics rather than second or third.

It is intriguing to note that the quasi-Bax output gives about the same HF THD as the emitter follower topology, confirming the statement that the addition of a Baxandall diode turns a conventional quasi-complementary

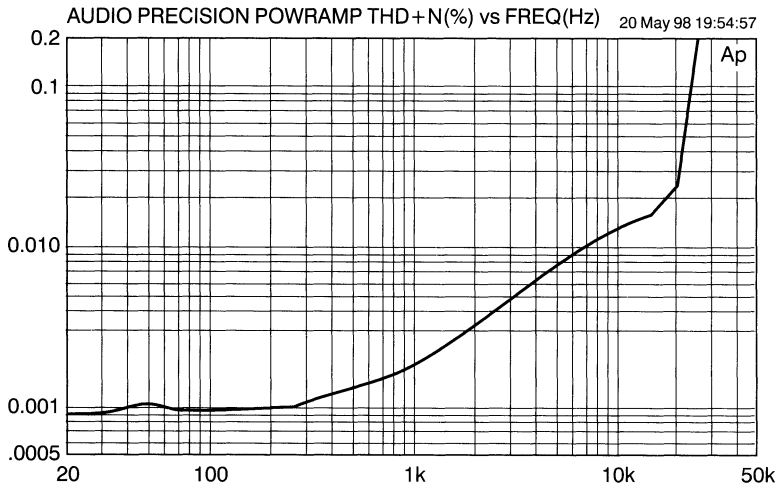



**Figure 5** *Closed-loop amplifier performance with emitter follower output stage.* 100 W *into* 8 Ω.

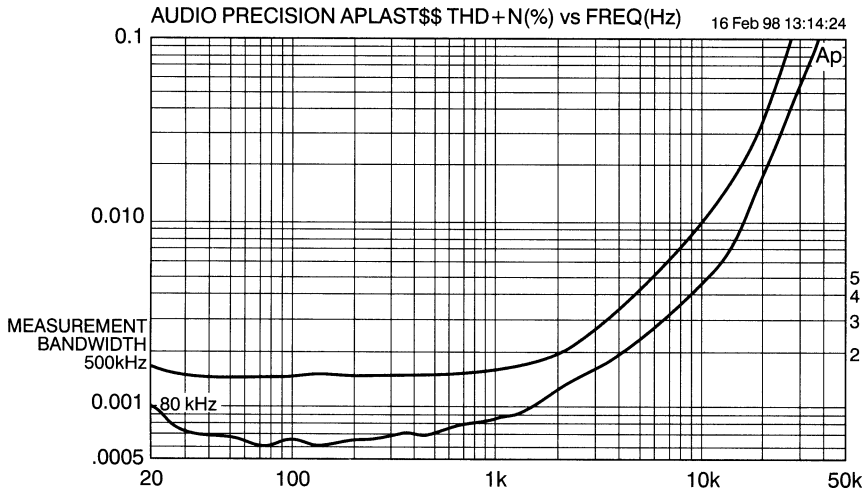



**Figure 6** *Closed-loop amplifier performance with CFP output.* 100 W *into* 8 Ω.

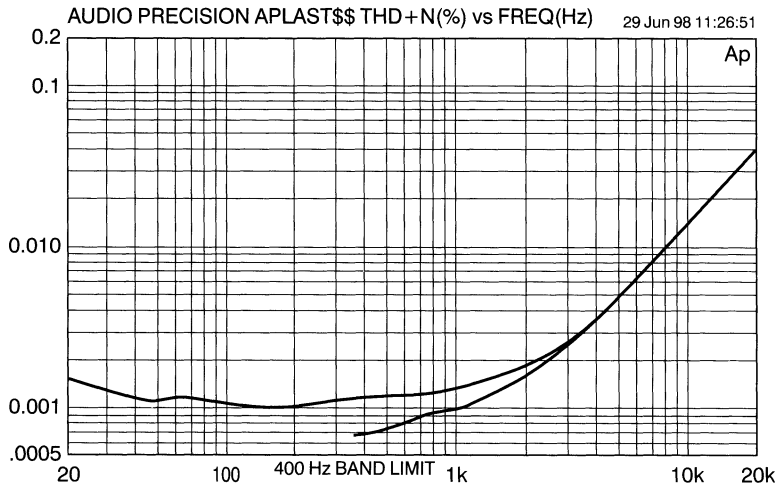



**Figure 7** *Closed-loop amplifier performance; quasi-complementary output stage with Baxandall diode.* 100 W *into* 8 Ω. *AP plots in Figures 5 to 7 were taken at* 100 Wrms/8 Ω, *from an amplifier with an input error of* −70 dB *at* 10 kHz *and c/1 gain of* 27 dB, *giving a feedback factor of* 43 dB *at this frequency. This is well above the dominant pole frequency, so the NFB factor is dropping at* 6 dB/*octave and will be down to* 37 dB *(or* 70×*) at* 20 kHz. *My experience suggests that this is about as much NFB as is safe for general use, assuming an output inductor to improve stability with capacitive loads. Sadly, published data on this touchy topic seems non-existent.*

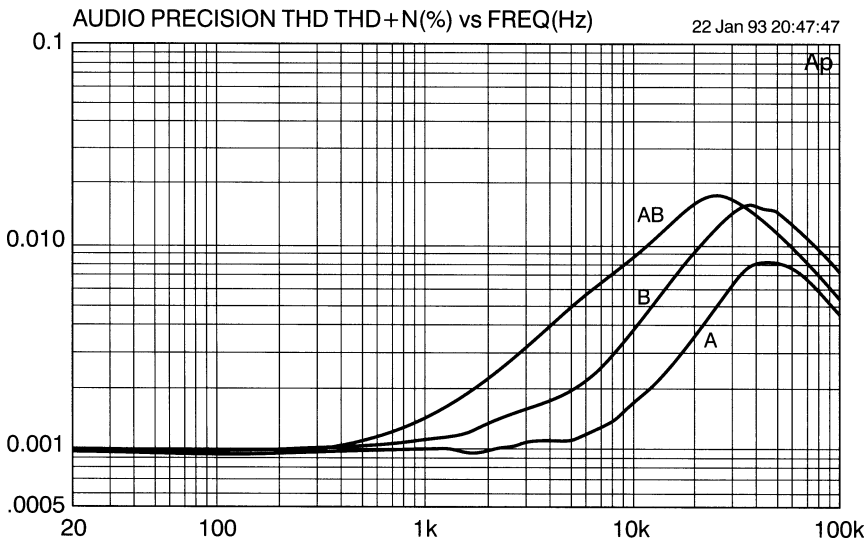



**Figure 8** *Closed-loop CFP amp. Setting quiescent for Class AB gives more HF THD than either Class A or B.*

stage with serious crossover asymmetry into a reasonable emulation of a complementary emitter follower stage.

There is significantly less HF THD with a CFP output; this cannot be due to large signal nonlinearity as this is negligible with an 8 Ω load for all three stages, and must result from lower levels of high order crossover products.

Despite the promising ingredients, a distortionless amplifier has failed to materialise, so we had better find out why?

When an amplifier with a frequency dependent NFB factor produces distortion, the reduction is not due to the NFB factor at the fundamental frequency, but the amount available at the frequency of the harmonic in question.

## Harmonic generation by crossover distortion

The usual nonlinear distortions generate most of their unwanted energy in low order harmonics that NFB can deal with effectively. However, crossover and switching distortions that warp only a small part of the output swing tend to push energy into high order harmonics, and this important process is demonstrated here, by Fourier analysis of a Spice waveform.

Take a sinewave fundamental, and treat the distortion as an added error signal $E$, letting the ratio $WR$ describe the proportion of the

cycle where $E > 0$. If this error is a triangle wave extending over the whole cycle ($WR = 1$) this would represent large signal nonlinearity, and Figure 9 shows that most of the harmonic energy goes into the third and fifth harmonics; the even harmonics are all zero due to the symmetry of the waveform.

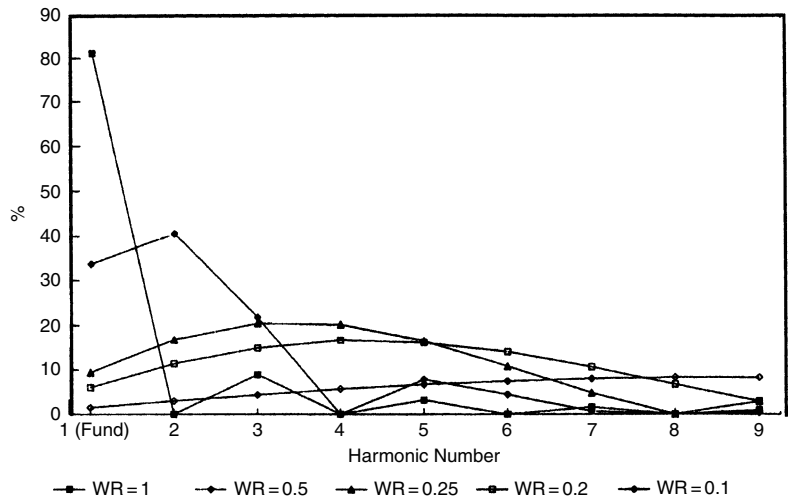



**Figure 9** *The amplitude of each harmonic changes with* WR*; as the error waveform gets narrower, energy is transferred to the higher harmonics.*

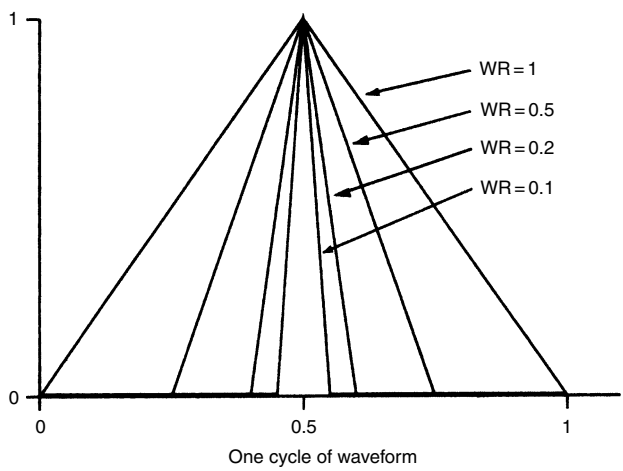



**Figure 10** *Diagram of the error waveform* E *for some values of* WR.

Figure 10 shows how the situation is made more like crossover or switching distortion by squeezing the triangular error into the centre of the cycle so that its value is zero elsewhere; now $E > 0$ for only half the cycle (denoted by $WR = 0.5$) and Figure 9 shows that the even harmonics are no longer absent. As *WR* is further decreased, the energy is pushed into higher order harmonics, the amplitude of the lower harmonics falling.

These high harmonics have roughly equal amplitude, spectrum analysis confirming that even in a blameless amplifier driven at 1 kHz, harmonics are freely generated from the 7th to the 19th at a level within a dB or so. The 19th harmonic is only 10 dB below the third.

Thus, in an amplifier with crossover distortion, the order of the harmonics will decrease as signal amplitude reduces, and *WR* increases; their lower frequencies allow them to be better corrected by the frequency dependant negative feedback. This effect seems to work against the commonly assumed rise of percentage crossover distortion as level is reduced.

A typical amplifier with open loop gain rolling off at 6 dB/octave will be half as effective at reducing fourth-harmonic distortion as it is at reducing the second harmonic. LSN is largely third (and possibly second) harmonic, and so NFB will deal with this effectively. However, both crossover and switchoff distortions generate high-order harmonics significant up to at least the 19th and these receive much less linearisation. As the fundamental moves up in frequency the harmonics do too, and get even less feedback. This is the reason for the differentiated look to many distortion residuals; higher harmonics are emphasised at the rate of 6 dB/octave.

Here is a real example of the inability of NFB to cure all amplifier ills. To reduce this HF distortion we must reduce the crossover gain deviations of the output stage before closing the loop. There seems no obvious way to do this by minor modifications to any of the conventional output stages; we can only optimise the quiescent current.

Increasing the quiescent current will do no good for, as outlined in the previous chapter, Class AB is generally Not A Good Thing, producing more distortion than Class B, not less. Figure 8 makes this painfully clear for the closed-loop case; Class AB clearly gives the worst performance. (As before, the AB quiescent was set for 50:50 m/s ratio of the $g_m$ doubling artifacts on the residual).

In this case the closed loop distortion is much greater than that from the small signal stages alone; however this is not automatic, and if the input pair is badly designed its HF distortion can easily exceed that caused by the output stage.

The distortion figures given in this article are rather better than the usual run. I must emphasise that these are not freakish or unrepeatable figures. They are simply the result of attending to all seven of the major sources of distortion rather than just one or two. I have so far built 12 CFP amplifiers, and performance shows little variation.

### *Conclusions*

Taking this and the previous chapter together, we can summarise. Class AB is best avoided. Use pure Class A or B, as AB will always have more distortion than either. Fet outputs offer freedom from some BJT problems, but in general have poorer linearity, lower efficiency, and cost more.

Distortion generated by a blameless amplifier driving an 8Ω load is almost entirely due to crossover effects and switching distortion. This does not hold for 4Ω or lower loads where third harmonic on the residual shows the presence of large signal nonlinearity caused by beta loss at high output currents.

## References

1. Gray and Meyer *Analysis & Design of Analog Integrated Circuits*, Wiley 1984, p 172.
2. Crecraft *et al. Electronics*, Chapman & Hall 1993, p 538.
3. Oliver 'Distortion in complementary pair class B amps', *Hewlett Packard Journal*, February 1971, p 11.
4. Blomley, P. 'New approach to class B', *Wireless World*, February 1971, p 57.
5. Thus, F.A. 'Compact bipolar class AB output stage using 1V power supply', *IEEE Journal of Solid State Circuits*, December 1992, p 1718 (Harmonic AB).
6. Tanaka, S.A. 'New biasing circuit for class B operation', *JAES*, January/February 1981, p 27.
7. Hawksford, M. 'Distortion correction in audio power amplifiers', *JAES*, March 1981, p 148.
8. Alves, J. 'Power bandwidth limitations in audio amplifiers', *IEEE Trans on Broadcast & TV*, March 1973, p 79.
9. Baxandall, P. 'Symmetry in class B', *Letters, Wireless World*, September 1969, p 416 (Baxandall diode).

# 21 Distortion in power amplifiers, Part VI: the remaining distortions

*January 1994*

The previous two chapters considered closely the distortion produced by amplifier output stages: a basically conventional but well designed Class-B amplifier with proper precautions taken against the various sources of nonlinearity can produce insignificant levels of distortion. That which is generated is mainly due to the difficulty of reducing high order crossover nonlinearities with negative feedback that has declining effectiveness with frequency. For 8 Ω loads this is the major source of distortion. For convenience, I have chosen to call such a device a *blameless* amplifier.

## Distortion 3: quiescent current control

An optimised amplifier requires minimisation of output stage gain irregularities around the crossover point by holding the quiescent current $I_q$ at its optimal value. Increasing $I_q$ to move into Class-AB makes the distortion worse, not better, as $g_m$-doubling artifacts are generated.

The initial setting of quiescent current is simple, given a distortion analyser to get a good view of the residual; keeping that setting under varying operating conditions is a much greater problem because $I_q$ depends on small voltages established across low value resistors by power devices with thermally dependant $V_{be}$ drops.

How accurately does quiescent current need to be maintained? I wish I could be more specific on this. Some informal experiments with Blameless CFP type outputs at 1 kHz indicate that crossover artifacts on THD residual seem to stay at roughly the same level, partly submerged in the noise, over

## Blameless amplifiers

I have adopted the term *blameless* to describe a Class-B amplifier designed in accordance with the philosophy of this series, with the use of simple circuit enhancements to minimise distortions 1,2 and 4, and correct layout to prevent distortions 5,6 and 7. Such a device will still suffer from output stage distortion 3, and so exhibit measurable distortion at high frequencies due to the difficulty that NFB has in dealing with the high order crossover distortion products generated by a conventional (but well designed) output stage. Distortion will usually be greater when driving loads below 8 Ω.

The word is specifically chosen to imply the avoidance of error but not perfection.

an $I_q$ range of about 2:1, the centre of this region being around 20 mA. Results may well be different for emitter follower type outputs.

This may seem a wide enough target, but given that junction temperature of power devices may vary over a 100 °C range, this is not so. Some kinds of amplifier (e.g. current dumping types) manage to evade the problem altogether, but in general the solution is thermal compensation: the output stage bias voltage is set by a temperature sensor (usually a $V_{be}$ multiplier transistor) coupled as closely as possible to the power devices.

There are inherent inaccuracies and thermal lags in this sort of arrangement leading to programme dependency of $I_q$. A sudden period of high power dissipation will begin with the bias current increasing above optimum, as the junctions will heat up very quickly. Eventually the thermal mass of the heatsink will respond, and the bias voltage will be reduced. When the power dissipation falls again, the bias voltage will now be too low to match the cooling junctions and the amplifier will be under biased, producing crossover spikes that may persist for some minutes. This is well illustrated in an important paper by Sato.[1]

## Emitter follower outputs

The major drawback of emitter follower output stages is thermal stabilisation. This can cause production problems in initial setting up since any drift of quiescent current will be very slow as a lot of metal must warm up.

For EF outputs, the bias generator must attempt to establish an output bias voltage that is a summation of four driver and output $V_{be}$'s. These do not vary in the same way. It seems at first a bit of a mystery how the EF

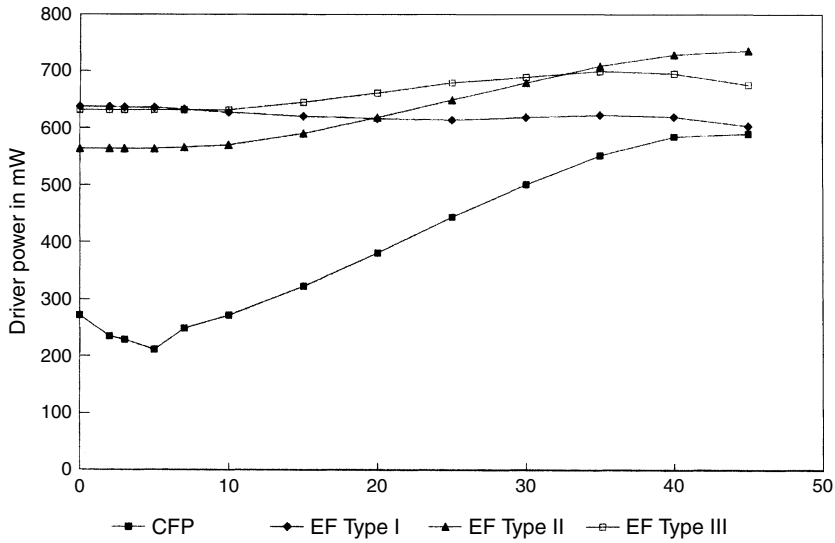



**Figure 1** *The variation in driver dissipation with output for the three EF output topologies and the CFP output. All three EF types keep driver power fairly constant, simplifying the thermal compensation problem.*

stage, which still seems to be the most popular output topology, works as well as it does. The probable answer is Figure 1, which shows how driver dissipation (averaged over a complete cycle) varies with peak output level for the three kinds of EF output, and for the CFP configuration. The Spice simulations used to generate this graph used a triangle waveform to give a slightly closer approximation to the peak-average ratio of real waveforms. The rails were $\pm 50\,V$, and the load $8\,\Omega$.

It is clear that the driver dissipation for the EF types is relatively constant with power output, while the CFP driver dissipation, although generally lower, varies strongly. This is a consequence of the different operation of these two kinds of output. In general, the drivers of an EF output remain conducting to some degree for most or all of a cycle, although the output devices are certainly off half the time.

In the CFP, however, the drivers turn off almost in synchrony with the outputs, dissipating an amount of power that varies much more with output. This implies that EF drivers will work at roughly the same temperature, and can be neglected in arranging thermal compensation; the temperature dependent element is usually attached to the heatsink to compensate for the junction temperature of the output devices alone. The Type I EF output keeps its drivers at the most constant temperature.

The above does not apply to integrated Darlington outputs, with drivers and assorted emitter resistors combined in one ill-conceived package where

the driver sections are directly heated by the output junctions. This works directly against quiescent stability.

The drawback with most thermal compensation schemes is the slow response of the heatsink mass to thermal transients. The obvious solution is to find some way of getting the sensor closer to one of the output junctions. If TO3 devices are used, then the flange on which the actual transistor is mounted is as close as one can get without a hacksaw. This is however clamped to the heatsink, and almost inaccessible, though it might be possible to hold a sensor under one of the mounting bolts. A simpler solution is to mount the sensor on the top of the TO3 can. This is probably not as accurate an estimate of junction temperature as the flange would give, but measurement shows the top gets much hotter much faster than the heatsink mass, so while it may appear unconventional, it is probably the best sensor position for an EF output stage.

Figure 2 shows the results of an experiment designed to test this. A TO3 device was mounted on a thick aluminium L-section thermal coupler in turn clamped to a heatsink; this construction represents many typical designs. Dissipation equivalent to 100 W/8 Ω was suddenly initiated, and the temperature of the various parts monitored with thermocouples. The graph clearly shows that the top of the TO3 responds much faster, and with a larger temperature change, though after the first two minutes the temperatures are all increasing at the same rate. The whole assembly took more than an hour to asymptote to thermal equilibrium.

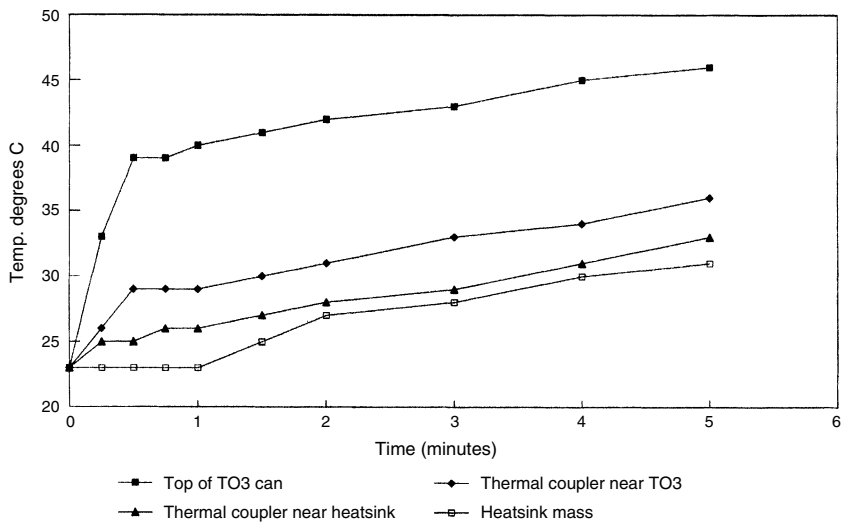



**Figure 2** *Thermal response of a TO3 coupled to a large heatsink when power is abruptly applied. The top of the TO3 can responds most rapidly.*

## The CFP output

In the CFP configuration, the output devices are inside a local feedback loop, and play no significant part in setting $I_q$, which is affected only by thermal changes in the drivers' $V_{be}$. Such stages are virtually immune to thermal runaway; I have found that assaulting the output devices with a powerful heat gun induces only insignificant $I_q$ changes. Thermal compensation is mechanically simpler as the $V_{be}$ multiplier transistor is simply mounted on one of the driver heatsinks, where it aspires to mimic the driver junction temperature. It is now practical to make the bias transistor of the same type as the drivers, which should give the best matching of $V_{be}$, though how important this is in practice I wouldn't like to say.[2]

Because driver heatsinks are much smaller than the main heatsink, the thermal compensation time constant is now measured in tens of seconds rather than tens of minutes, and should give much shorter periods of non optimal quiescent current than the EF output topology.

## Distortion 4: nonlinear loading of the voltage amplifier stage by the nonlinear impedance of the output stage

This distortion mechanism was examined in Chapter 18 from the point of view of the voltage amplifier stage (VAS). Essentially, since the VAS provides all the voltage gain, its collector impedance tends to be made high. This renders it vulnerable to nonlinear loading unless it is buffered.

Making a linear VAS is most easily done by applying a healthy amount of local negative feedback via the dominant pole Miller capacitor, and if VAS distortion needs further reduction, then the open loop gain of the VAS stage must be raised to increase this local feedback. The direct connection of a Class-B output can make this difficult for, if the gain increase is attempted by cascoding with intent to raise the impedance at the VAS collector, the output stage loading will render this almost completely ineffective. The use of a VAS buffer eliminates this effect.

As explained previously, the collector impedance, while high at LF compared with other circuit nodes, falls with frequency as soon as $C_{dom}$ starts to take effect, and so the fourth distortion mechanism is usually only visible at LF. It is also masked by the increase in output stage distortion above dominant pole frequency P1 as the amount of global NFB reduces.

The fall in VAS impedance with frequency is demonstrated in Figure 3, obtained from the Spice conceptual model outlined previously, with real life values. The LF impedance is basically that of the VAS collector resistance, but halves with each octave once P1 is reached. By 3 kHz it is down

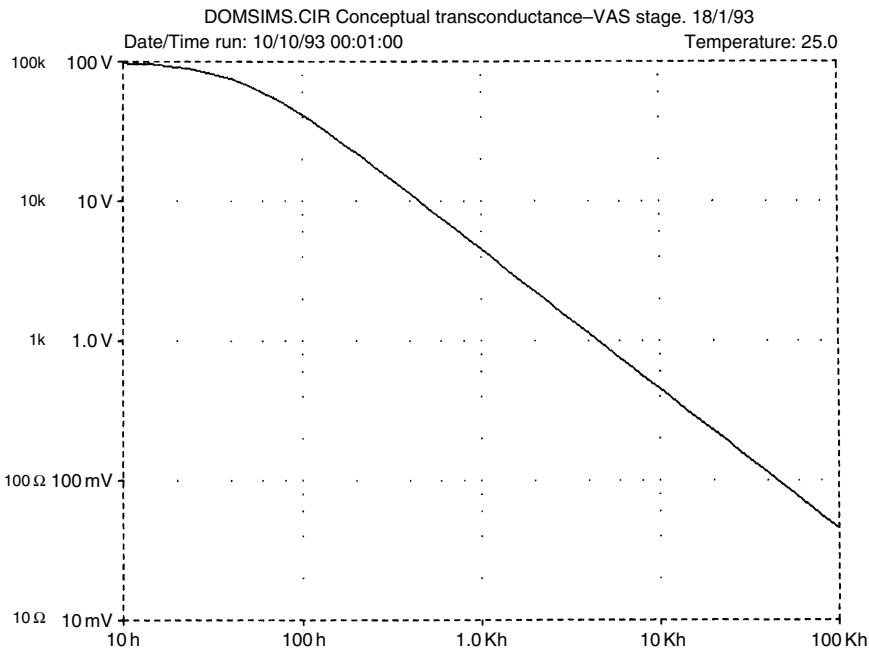



**Figure 3** *Distortion 4. The impedance at the VAS collector falls at* 6 dB/octave *with frequency.*

to 1 kΩ and still falling. Nevertheless, it can remain high enough for the input impedance of a Class-B output stage to significantly degrade linearity, the actual effect being shown in Figure 4.

An alternative to cascoding for VAS linearisation is to add an emitter follower within the VAS local feedback loop, increasing the local NFB factor by raising effective beta rather than the collector impedance. Preliminary tests show that as well as providing good VAS linearity, it establishes a lower VAS collector impedance across the audio band. It should be more resistant to this type of distortion than the cascode version.

Figure 5 confirms that the input impedance of a conventional EF Type I output stage is anything but linear; the data is derived from a Spice output stage simulation with optimal $I_q$. Even with an undemanding 8 Ω load, the impedance varies by 10:1 over the output voltage swing. Interestingly, the Type II EF output (using a shared drive emitter resistance) has a 50% higher impedance around crossover, but the variation ratio is rather greater. CFP output stages have a more complex variation that includes a precipitous drop to less than 20 kΩ around the crossover point. With all types under biasing produces additional sharp impedance changes at crossover.

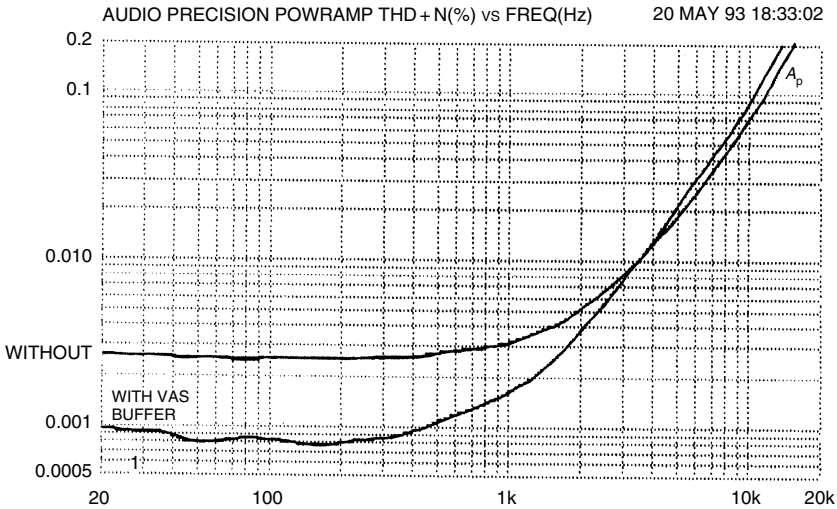



**Figure 4** *Distortion 4 in action. The lower trace shows the result of its elimination by the use of a VAS buffer.*

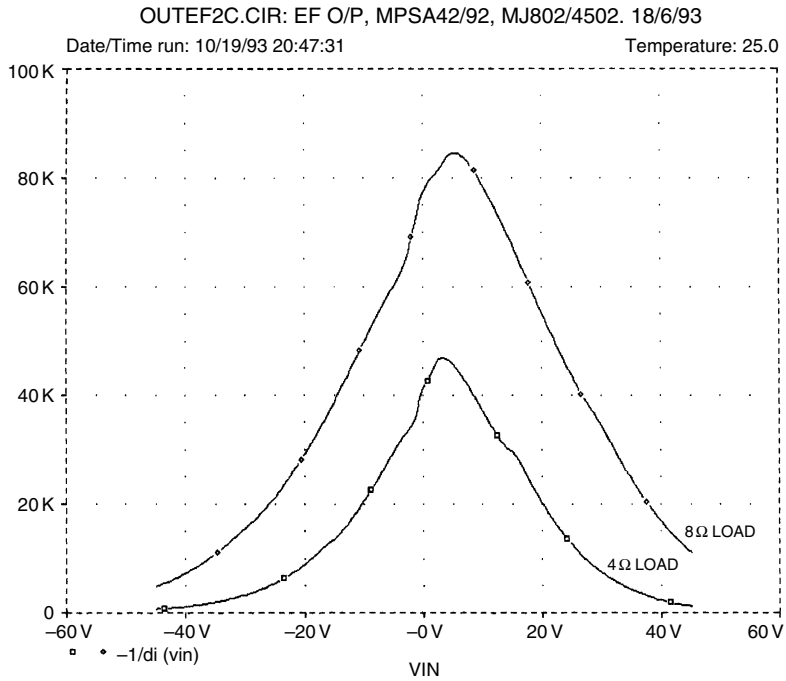



**Figure 5** *Distortion 4 and its root cause. The nonlinear input impedance of an EF Class B output stage.*

## Distortion 5: supply ground loops

Virtually all amplifiers include some form of rail decoupling apart from the main reservoir capacitors; this is usually required to improve HF stability. The standard decoupling arrangements include small to medium sized electrolytics (say 10–1000 μF) connected between each rail and ground, and an inevitable consequence is that voltage variations on the rails cause current to flow into the ground connection chosen. This is just one mechanism that defines the power supply rejection ratio (PSRR) of an amplifier, but it is one that can do serious damage to linearity. If we assume a simple unregulated power supply, (and there are excellent reasons for using such a supply[3]) then these rails have a significant A.C. impedance and superimposed voltage will be due to amplifier load currents as well as 100 Hz ripple. In Class-B, these supply rail currents are halfwave rectified sine pulses with strong harmonic content, and if they contaminate the signal, then distortion will degrade badly. A common route for interaction is via decoupling grounds shared with input or feedback networks, and a completely separate decoupler ground usually effects a total cure. This point is easy to overlook, and attempts to improve amplifier linearity by labouring on the input pair, VAS, etc., are doomed to failure unless this distortion mechanism is eliminated first.

As a rule it is simply necessary to take the decoupling ground separately back to the ground star point, as shown in Figure 6. Note that the star point A is defined on a short spur from the heavy connection joining the reservoirs; trying to use B as the star point will introduce ripple due to the large reservoir charging current pulses passing through it.

Figure 7 shows the effect on an otherwise optimised amplifier delivering 60 W/8 Ω with 220 μF rail decoupling capacitors. At 1 kHz distortion has increased by more than ten times, which is quite bad enough. However, at

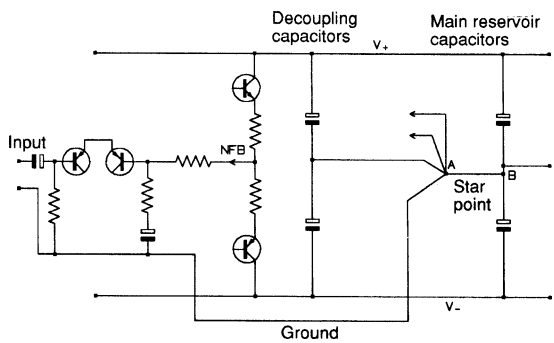



**Figure 6** *Distortion 5. The correct way to route decouple grounding to the star point.*

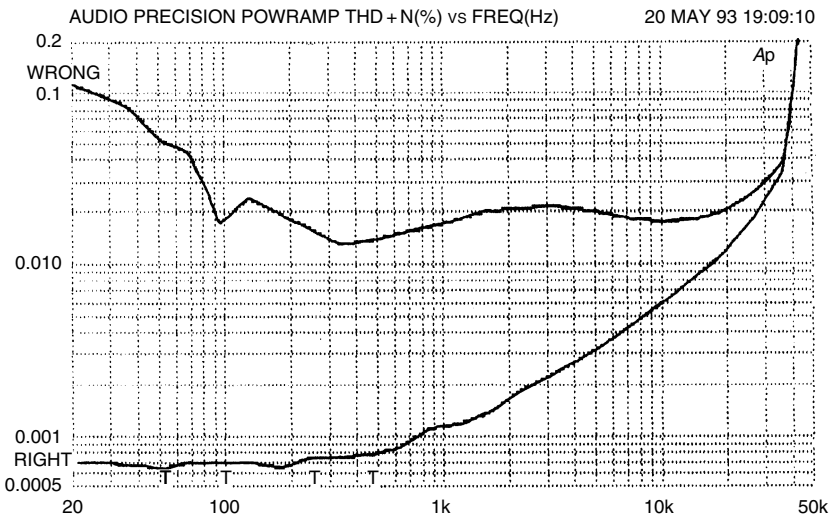



**Figure 7** *Distortion 5 in action. The upper trace was produced simply by taking the decoupler ground from the star point and connecting it via the input ground line instead.*

20 Hz the THD has increased at least 100 fold, turning a very good amplifier into a profoundly mediocre one with a single misconceived connection.

If the residual on the supply rails is examined, the ripple amplitude will usually be found to exceed the pulses due to Class-B signal current, and so some of the 'distortion' on the upper curve of the plot is actually due to ripple injection. This is hinted at by the phase crevasse at 100 Hz, where ripple partly cancelled the signal at the instant of measurement. Below 100 Hz the – curve rises as greater demands are made on the reservoirs, the signal voltage on the rails increases, and so more distorted current is forced into the ground system.

Generally, if an amplifier is made free from ripple injection under drive conditions, shown by a THD residual without ripple components, there will be no distortion from the supply rails and the complications and inefficiency of high current rail regulators are unnecessary.

There has been much discussion of PSRR induced distortion in *EW + WW* recently, led by Ben Duncan[4] and Greg Ball.[5] I part company with Ben Duncan on this issue where he assumes that a power amplifier is likely to have 25 dB PSRR, making expensive high power DC regulators the only answer. He agrees that this sort of PSRR is highly unlikely with the relatively conventional amplifier topologies I have been considering.[6]

Greg Ball also initially assumes that a power amp has the same PSRR characteristics as an op-amp, i.e. falling steadily at 6 dB/octave. There is absolutely no need for this to be so, given a little RC decoupling, and

Ball states at the end of his article that 'a more elegant solution . . . is to depend on a high PSRR in the amplifier proper.'

## Power supply rejection

For low noise and distortion, all the obvious methods of rail injection must be attended to as a matter of routine. I therefore give here some guidelines that I have found effective with unregulated supplies:

- The input pair must have a tail current source. A tail made of two resistors decoupled mid way is simply not adequate.
- This tail source will probably be biased by a pair of diodes or a led fed from a resistor to ground. This resistor should be split and the midpoint decoupled with an electrolytic of about 10 μF to the appropriate rail.
- If a cascode transistor is used in the VAS, then its base will need to be biased about 1.2 V above whichever rail the VAS emitter sits on; if this is implemented with a pair of diodes then further decoupling seems unnecessary.
- Having taken care of the above, the PSRR will now be limited by injection from the negative rail by a mechanism that is not yet fully clear. RC decoupling can however reduce this to negligible levels.

This is not the whole story on power rail rejection, but it does provide a starting point.

## Distortion 6: induced output current coupling

This distortion mechanism, like the previous case, stems directly from the Class-B nature of the output stage. Assuming a sine input, the output hopefully carries a good sinewave, but the supply rail currents are halfwave rectified sine pulses, which are quite capable of inductive crosstalk into sensitive parts of the circuit. This can be very damaging to the distortion performance, as Figure 8 shows.

The distortion signal may intrude into the input circuitry, the feedback path, or even the cables to the output terminals. The result is a kind of sawtooth on the distortion residual that is very distinctive, an extra distortion component which rises at 6 dB/octave with frequency.

This effect appears to have been first publicised by Cherry,[7] in a paper that deserves much more attention than it appears to have got. Having examined many power amplifiers, I feel that this effect is probably the most widespread cause of unnecessary distortion.

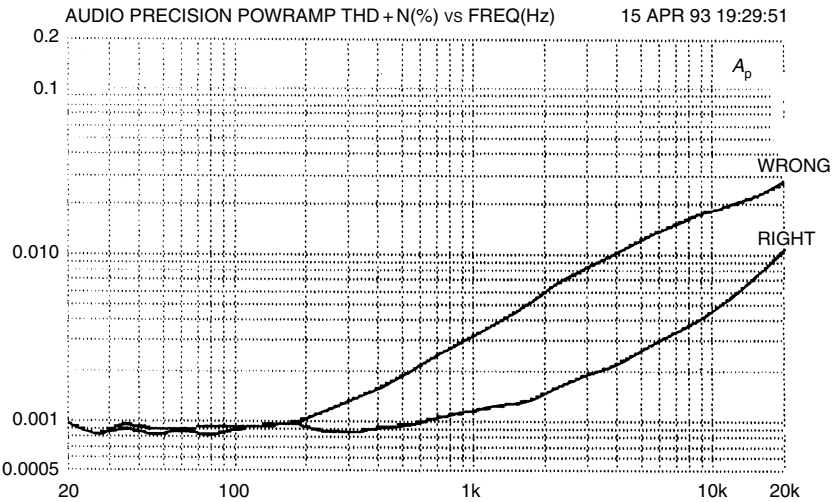



**Figure 8** *Distortion 6 exposed. The upper trace shows the effects of Class B rail induction into signal circuitry.*

Effects of this distortion mechanism can be reduced below the measurement threshold by taking care over supply rail cabling layout relative to signal leads, and avoiding loops that will induce or pick up magnetic fields. There are no precise rules for layout that would guarantee freedom from rail induction since each amplifier has its own physical layout and the cabling topology needs to take this into account. All I can do is give guidelines:

- Firstly, implement rigorous minimisation of loop area in the input and feedback circuitry; keep each signal line as close to its ground return as possible.
- Secondly, minimise the ability of the supply wiring to create magnetic fields.
- Thirdly, put as much distance between these two areas as you can. Fresh air beats shielding.

on price every time. Figure 9(a) shows one straightforward approach to tackling the problem; the supply and ground wires are tightly twisted together to reduce radiation. In practice this doesn't seem too effective for reasons that are not wholly clear, but appear to involve the difficulty of ensuring exactly equal coupling between three twisted conductors.

In Figure 9(b), the supply rails are twisted together but kept well away from the ground return. This allows field generation, but if currents in the two rails butt together to make a sinewave at the output, they should do the same when the magnetic fields from each rail sum. There is an obvious

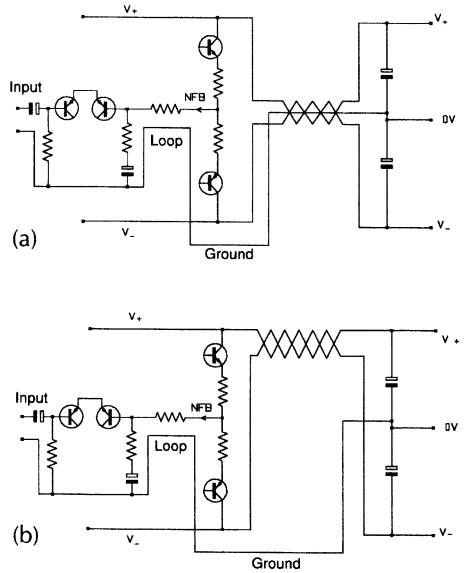



**Figure 9** *Distortion 6. Countermeasures against the induction of distortion from the supply rails. 9(b) is usually more effective.*

risk of interchannel crosstalk with this approach in a stereo amplifier, but it does seem to deal most effectively with the induced distortion problem.

## Distortion 7: nonlinearity from incorrect NFB connection point

Negative feedback is a powerful technique and must be used with care. Designers are repeatedly told that too much feedback can affect slew rate. Possibly true, though the greater danger is that an excess amplifier may produce tweeter-frying HF instability.

However, there is another and more subtle danger. Class-B output stages are a hotbed of high amplitude halfwave rectified currents, and if the feedback takeoff point is even slightly asymmetric, these will contaminate the feedback signal making it an inaccurate representation of the output voltage. This will manifest itself as distortion, Figure 10.

At the current levels in question, all wires and PCB tracks must be treated as resistances, and it follows that point *C* is not at the same potential as point *D* whenever $TR_1$ conducts. If feedback is taken from *D*, then a clean signal will be established here, but the signal at output point *C* will have a half wave rectified sinewave added to it, due to the resistance *C–D*. The

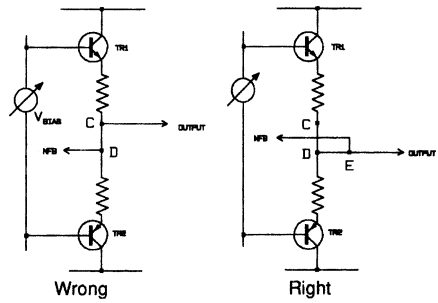



**Figure 10** *Distortion 7. Wrong and right ways of arranging the critical negative feedback takeoff point.*

output will be distorted but the feedback loop will do nothing about it as it does not know about the error.

Figure 11 shows the practical result for an amplifier driving 100 W into 8 Ω, with the extra distortion shadowing the original curve as it rises with frequency. Resistive path *C–D* that did the damage was a mere 6 mm length of heavy gauge wirewound resistor lead.

Elimination of this distortion is easy, once you know the danger. Connecting the feedback arm to *D* is not advisable as it will not be a mathematical point, but will have a physical extent inside which the current distribution is unknown. Point *E* on the output line is much better, as the half wave currents do not flow through this arm of the circuit.

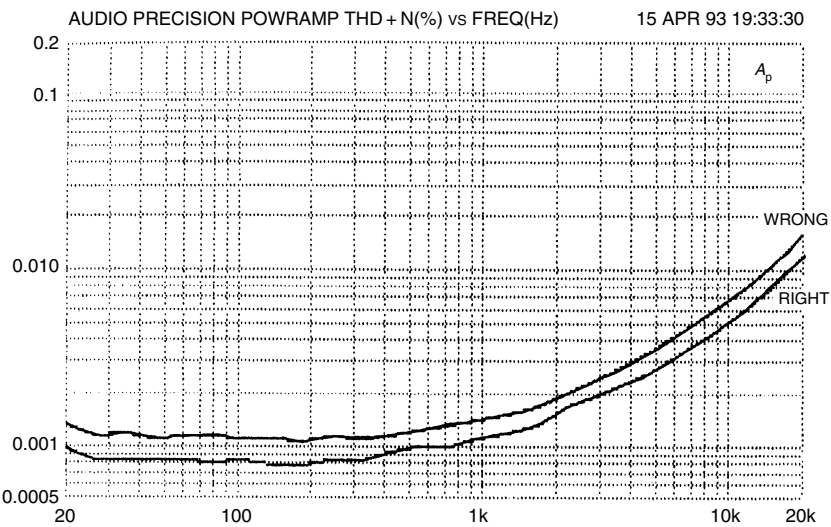



**Figure 11** *Distortion 7 at work. The upper trace shows the result of a mere* 6 mm *of heavy gauge wire between the output and the feedback point.*

## References

1. Sato 'Amplifier transiant grossover distortion'. AES Preprint for 72nd Convention, October 1982.
2. Evans, J 'Audio amplifier bias current', Letters, *Electronics + Wireless World*, January 1991, p 53.
3. Self, D 'Sound mosfet design', *Electronics + Wireless World*, September 1990, p 760.
4. Duncan, B 'PSU regulation boosts audio performance', *Electronics + Wireless World*, October 92, p 818.
5. Ball, G 'Distorting power supplies', *Electronics + Wireless World*, December 1990, p 1084.
6. Duncan, B, Private communication, October 93.
7. Cherry 'A New distortion mechanism in class-B amplifiers', *JAES* May 1981, p 327.

# 22 Distortion in power amplifiers, Part VII: frequency compensation and real designs

*February 1994*

The distortion performance of an amplifier is determined not only by open loop linearity, but also the negative feedback factor applied when the loop is closed. In most practical circumstances doubling the NFB factor halves the distortion. To date, this series has focused on basic circuit linearity. I have assumed that open loop gain falls at 6 dB/octave due to a single dominant pole, with the amount of NFB permissible at HF being set by the demands of HF stability. Because of this, the distortion residuals from a 'blameless' amplifier are comprised almost entirely of crossover artifacts due to their high frequency content. Audio amplifiers using more advanced compensation are rather rare. However, certain techniques do exist . . .

This series has stuck close to conventional topologies, because even commonplace circuitry has been shown to have little known aspects and interesting possibilities. This implies a two-gain-stage circuit (unity gain output stages not being counted) with most of the feedback applied globally, but smoothly transferred to the voltage amplifier stage alone as frequency increases. Other configurations are possible; a one stage amplifier is an intriguing possibility – they are common in cmos op-amps – but is probably ill-suited to power amp impedances. See Ref. 1 for an eccentric three-stage amplifier with an open loop gain of just 52 dB (due to the dogged use of local feedback) and only 20 dB of global feedback. Most of the section below refers only to the conventional two-stage structure.

## Making a pole dominant

Dominant pole compensation is the simplest kind, though its implementation involves subtlety. Simply take the lowest pole to hand (PI), and make it dominant, i.e. so much lower in frequency than the next pole P2 that the total loop gain (the open loop gain as reduced by the attenuation in the feedback network) falls below unity before enough phase shift accumulates to cause HF oscillation. With a single pole, the gain must fall at 6 dB/octave, corresponding to a constant 90° phase shift. Thus the phase margin will be 90° giving good stability. Figure 1(a) shows the traditional Miller method of making a dominant pole. The collector pole of $Tr_4$ is lowered by adding the Miller capacitance $C_{\mathrm{dom}}$ to that which unavoidably exists as the $C_{\mathrm{bc}}$ of the VAS transistor. However there are other beneficial

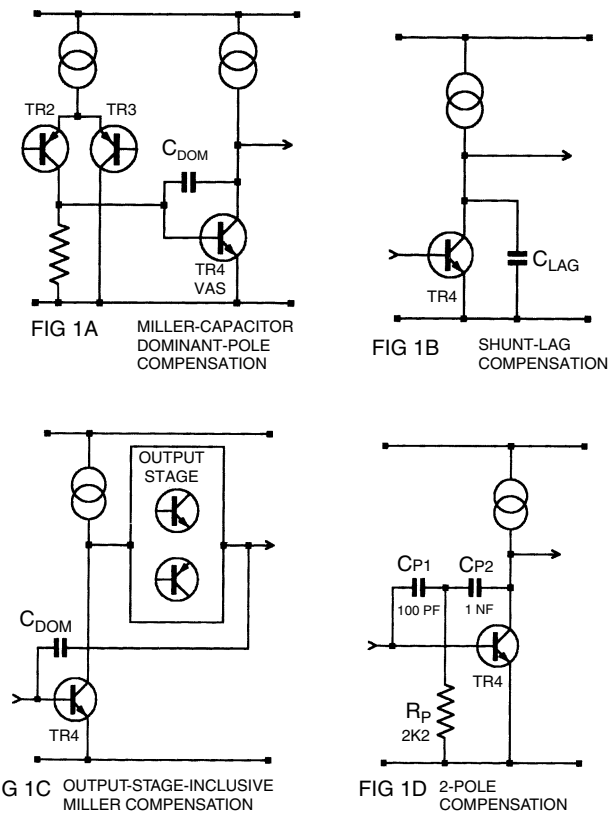



**Figure 1** *Implementing dominant-pole compensation. (a) Miller capacitor, (b) Shunt-lag circuit (c) Output-stage Inclusive Miller compensation. (d) How to implement 2-pole compensation.*

effects; $C_{dom}$ causes 'pole splitting', in which the pole at $Tr_2$ collector is pushed up in frequency as P1 moves down – most desirable for stability. Simultaneously the local NFB through $C_{dom}$ linearises the VAS.

Assuming that input stage transconductance is set to a plausible 5 mA/V, and stability considerations set the maximal 20 kHz open loop gain to 50 dB, then from the equations in Part 1, $C_{dom}$ must be 125 pF. This is more than enough to swamp the internal capacitances of the VAS transistor, and is a realistic value.

The peak current that flows in and out of this capacitor for an output of 20 V r.m.s., 20 kHz, is 447 μA. Recalling that the input stage must sink $C_{dom}$ current while the VAS collector load sources it, and likewise the input stage must source it while the VAS sinks it, there are four possible places in which slew rate might be limited by inadequate current capacity. If the input stage is properly designed then the usual limiting factor is VAS current sourcing. In this example a peak current of less than 0.5 mA should be easy to deal with, and the maximum frequency for unslewed output will be comfortably above 20 kHz.

Figure 1(b) shows a much less satisfactory method – the addition of capacitance to ground from the VAS collector. This is usually called shunt lag compensation, and as Peter Baxandall aptly put it, 'The technique is in all respects sub-optimal'.[2]

We have already seen in Part 3 that loading the VAS collector resistively to ground is a very poor option for reducing LF open loop gain, and a similar argument shows that capacitive loading to ground for compensation purposes is an even worse idea. To reduce open loop gain at 20 kHz to 50 dB as before, the shunt capacitor $C_{lag}$ must be 43.6 nF, which is a whole different order of things from 125 pF. The current flowing in $C_{lag}$ at 20 V r.m.s., 20 kHz, is 155 mA peak, which is going to require some serious electronics to provide it. This important result can be derived by simple calculation, and I have confirmed it with Spice simulation. The input stage no longer constrains the slew rate limits, which now depend entirely on the VAS.

A VAS working under these conditions is almost certain to have poor linearity. The current variations in the stage, caused by the extra loading, produces more distortion and there is now no local NFB through a Miller capacitor to correct it. To make matters worse, the dominant pole P1 will probably need to be set to a lower frequency than for the Miller case, to maintain the same stability margins, as there is now no pole splitting to raise the pole at the input stage collector. Hence $C_{lag}$ may have to be even larger, and require even higher peak currents. Takahashi has produced a fascinating paper on this approach,[3] showing one way of heaving about the enormous compensation currents required for good slew rates. The only thing missing is an explanation of why shunt compensation was chosen in the first place.

## Including the output stage

Miller capacitor compensation elegantly solves several problems at once, and the decision to use it is not hard. However the question of whether to include the output stage in the Miller feedback loop is less easy. Such inclusion (see Figure 1(c)) presents the desirable possibility that local feedback could linearise both the VAS and the output stage, with just the input stage left out in the cold as frequency rises and global NFB falls. This idea is most attractive as it would greatly increase the feedback available to linearise a Class B output stage.

There is certainly *some* truth in this where applying $C_{dom}$ around the output as well as the $V_{as}$ reduced the peak 1 kHz THD from 0.05% to 0.02%.[4] However, it should be pointed out that the output stage was deliberately under biased to induce crossover spikes because, with optimal bias, the improvement was too small to be either convincing or worthwhile. Also, this demonstration used a model amplifier with TO-92 'output' transistors. In my experience this technique just does not work with real power bipolars because it induces intractable HF oscillation.

The use of local NFB to linearise the VAS demands a tight loop with minimal extra phase shift beyond that inherent in the $C_{dom}$ dominant pole. It is permissible to insert a cascode or a small signal emitter follower into this local loop, but a sluggish output stage seems to be pushing the phase margin too far; the output stage poles are now included in the loop, which loses its dependable HF stability. Bob Widlar has stated that output stage behaviour must be well controlled up to 100 MHz for the technique to be reliable.[5] This would appear to be virtually impossible for discrete power stages with varying loads.

While I have so far not found 'Inclusive Miller compensation' to be workable myself, others may know different. If anyone can shed further light I would be most interested.

## Nested feedback loops

Nested feedback is a way to apply more NFB around the output stage without increasing the global feedback factor. The output has an extra voltage gain stage bolted on, and a local feedback loop is closed around these two stages. This NFB around the composite block reduces output stage distortion and increases frequency response, to make it safe to include in the global NFB loop.

Suppose that block $A_1$ (Figure 2(a)) is a distortionless small signal amplifier providing all the open loop gain and so including the dominant pole.

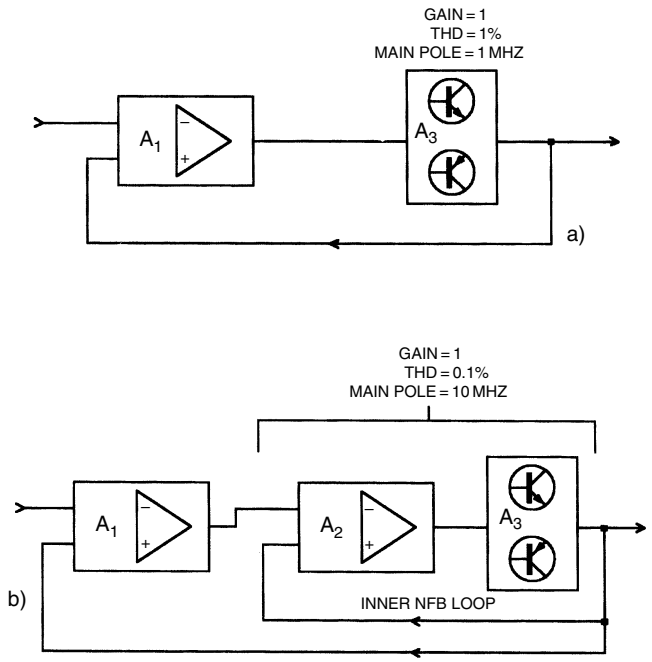



**Figure 2** *The principle of nested feedback loops.*

$A_3$ is a unity gain output stage with its own main pole at 1 MHz and distortion of 1% under given conditions: this 1 MHz pole puts a firm limit on the amount of global NFB that can be safely applied.

Figure 2(b) shows a nested feedback version; an extra gain block $A_2$ has been added, with local feedback around the output stage. $A_3$ has the modest gain of 20 dB so there is a good chance of stability when this loop is closed to bring the gain of $A_3 + A_2$ back to unity. $A_2$ now experiences 20 dB of NFB, bringing the distortion down to 0.1%, and raising the main pole to 10 MHz, which should allow the application of 20 dB more global NFB around the overall loop that includes $A_1$. We have thus decreased the distortion that exists before global NFB is applied, and simultaneously increased the amount of NFB that can be safely used, promising that the final linearity could be very good indeed. For another theoretical example see Ref. 6.

Real life examples of this technique in power amps are not easy to find, but a variation is widely used in op-amps. Many of us were long puzzled by the way that the much loved *5534* maintained such low THD up to high frequencies. Contemplation of its entrails appears to reveal a three-gain stage design with an inner Miller loop around the third stage, and an outer Miller loop around the second and third stages; global NFB is then applied

externally around the whole lot. Nested Miller compensation has reached its apotheosis in cmos op-amps – the present record appears to be three nested Miller loops plus the global NFB.[7] Don't try this one at home.

## Two pole compensation

Two pole compensation is a mildly obscure technique for squeezing the best performance from an op-amp,[8,9] but it has rarely been applied to power amplifiers. I know of only one example.[5] An extra HF time constant is inserted in the $C_{dom}$ path, giving an open loop gain curve that initially falls at almost 12 dB/octave, but which gradually reverts to 6 dB/octave as frequency continues to increase. This reversion is arranged to happen well before the unity loop gain line is reached, and so stability should be the same as for the conventional dominant pole scheme, but with increased negative feedback over part of the operational frequency range. The faster gain roll off means that the maximum amount of feedback can be maintained up to a higher frequency. There is no measurable mid band peak in the closed loop response.

One should be cautious about any circuit arrangement which increases the NFB factor. Power amplifiers face loads that vary widely: it is difficult to be sure that a design will always be stable under all circumstances. This makes designers rather conservative about compensation, and I approached this technique with some trepidation. However, results were excellent with no obvious reduction in stability. Figure 7 shows the result of applying this technique to the Class B amplifier described below.

The simplest way to implement two pole compensation is shown in Figure 1(d), with typical values. $C_{p1}$ should have the same value as it would for stable single pole compensation, and $C_{P2}$ should be at least twice as big; $R_P$ is usually in the region 1 k–10 k. At intermediate frequencies $C_{P2}$ has an impedance comparable with $R_p$, and the resulting extra time constant causes the local feedback around the VAS to increase more rapidly with frequency, reducing the open loop gain at almost 12 dB/octave.

At HF the impedance of $R_p$ is high compared with $C_{P2}$, the gain slope asymptotes back to 6 dB/octave, and then operation is the same as conventional dominant pole, with $C_{dom}$ equal to the series capacitance combination. So long as the slope returns to 6 dB/octave before the unity loop gain crossing occurs, there seems no obvious reason why the Nyquist stability should be impaired.

Figure 3 shows a simulated open loop gain plot for realistic component values; $C_{P2}$ should be at least twice $C_{P1}$ so the gain falls back to the 6 dB/octave line before the unity loop gain line is crossed. The potential feedback factor has been increased by more than 20 dB from 3 kHz to 30 kHz, a region where THD tends to increase due to falling

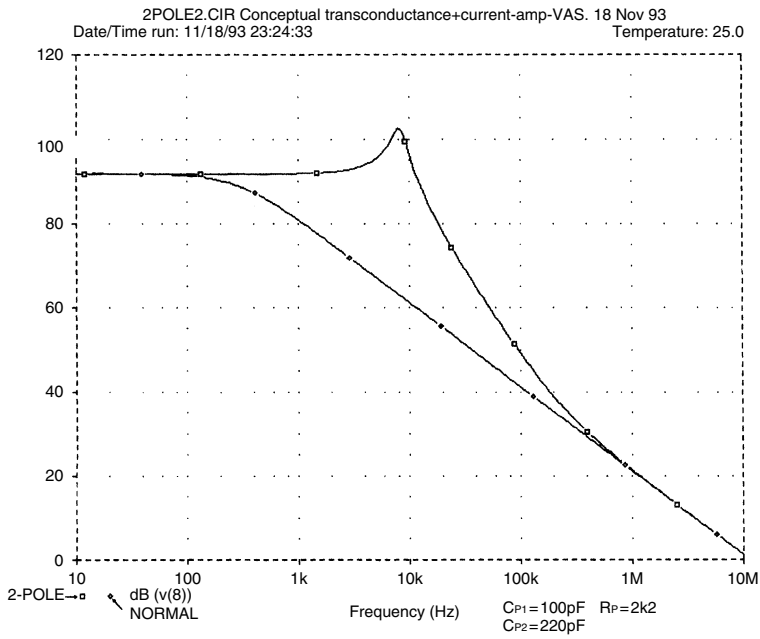



**Figure 3** *J3 Spice plot of the open-loop gain of a 2-pole compensated amplifier. The difference between the two plots shows the amount of extra NFB possible.*

NFB. The open loop gain peak at 8 kHz looks extremely dubious, but I have so far failed to detect any resulting ill effects in the closed loop behaviour.

There is however a snag to the simple approach shown here, which reduces the linearity improvement. Two-pole compensation may decrease open loop linearity at the same time as it raises the feedback factor that strives to correct it. At HF, $C_{P2}$ has low impedance and allows $R_p$ to directly load the VAS collector to ground. This worsens VAS linearity as we have seen. However, if $C_{P2}$ and $R_P$ are correctly proportioned the overall reduction in distortion is dramatic and extremely valuable. When two pole compensation was added to Figure 4, the crossover glitches on the THD residual almost disappeared, being partially replaced by low level 2nd harmonic which almost certainly results from VAS loading. The positive slew rate will also be slightly reduced.

This looks like an attractive technique, as it can be simply applied to an existing design by adding two inexpensive components. If $C_{P2}$ is much larger than $C_{P1}$, then adding/removing $R_p$ allows instant comparison between the two kinds of compensation. Be warned that if an amplifier is prone to HF parasitics then this kind of compensation may exacerbate them.

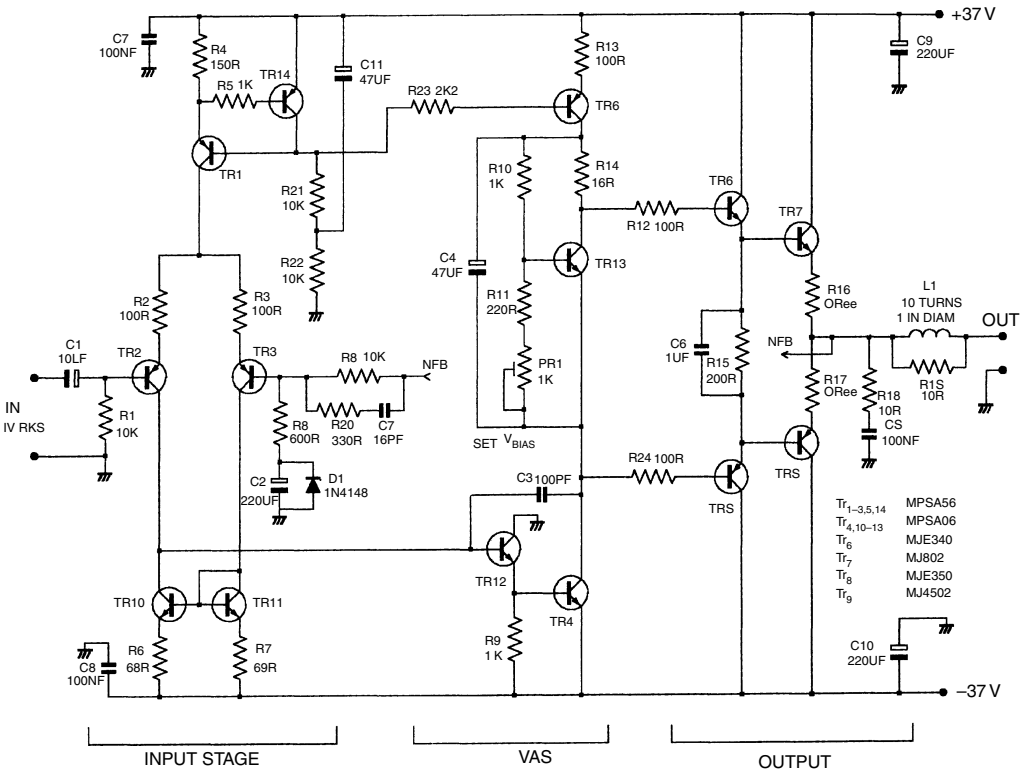



**Figure 4** 50 W *Class B amplifier circuit diagram. Transistor numbers correspond with the generic amplifier in chapter 16.*

## Design example: a 50 W class B amplifier

Figure 4 shows a design example of a Class B amplifier intended for domestic audio. Despite its conventional appearance, the circuit delivers a far better distortion performance than that normally associated with the arrangement.

With the supply voltages and values shown it delivers 50 W/8 Ω from IV r.m.s. input. In previous articles I have used the word *blameless* to describe amplifiers in which all distortion mechanisms, except the apparently unavoidable ones due to Class B, have been rendered negligible. This circuit has the potential to be *blameless*, but achieving this depends on care in cabling and layout. It does not aim to be a cookbook project; for example, overcurrent and DC offset protection are omitted.

The investigation presented in chapters 19 and 20 concluded that power fets were expensive, inefficient and non linear. Bipolars make good output devices. The best BJT configurations were the emitter follower type II, with least output switch-off distortion, and the complementary feedback pair (CFP) giving best basic linearity and quiescent stability.

I assume that domestic ambient temperatures will be benign, and the duty moderate, so that adequate quiescent stability can be attained by suitable heatsinking and thermal compensation. The configuration chosen is therefore emitter follower type II, which has the advantage of reducing switch-off nonlinearities (Distortion 3(c)) due to the action of $R_{15}$ in reverse biasing the output base emitter junctions as they turn off. The disadvantage is that quiescent stability is worse than for the CFP output topology, as there is no local feedback loop to servo out $V_{be}$ variations in the hot output devices.

The NFB factor was chosen as 30 dB at 20 kHz, which should give generous HF stability margins. The input stage (current source $Tr_1$, $Tr_{14}$ and differential pair $Tr_{2,3}$) is heavily degenerated by $R_2R_3$ to delay the onset of third harmonic Distortion 1. To assist this the contribution of transistor internal $r_e$ variation is minimised by using the unusually high tail current of 4 mA. $Tr_{10,11}$ form a degenerated current mirror that enforces accurate balance of the $Tr_{2,3}$ collector currents, preventing second harmonic distortion. Tail source $Tr_{1,14}$ has a basic PSRR 10 dB better than the usual two diode version, though this is academic when $C_{11}$ is fitted.

Input resistor $R_1$ and feedback arm $R_8$ are made equal and kept as low as possible consistent with a reasonably high input impedance, so that base current mismatch caused by beta variations will give a minimal DC offset. This does not affect $Tr_2$-$Tr_3$ $V_{be}$ mismatches, which appear directly at the output, but these are much smaller than the effects of $I_b$. Even if $Tr_{2,3}$ are high voltage types with low beta, the output offset should be within ±50 mV, which should be quite adequate, and eliminates balance presets

and DC servos. A low value for $R_8$ also gives a low value for $R_9$, which improves the noise performance.

The value of $C_2$ shown (220 µF) gives an LF roll off with $R_9$ that is −3 dB at 1.4 Hz. The aim is not an unreasonably extended sub-bass response, but to prevent an LF rise in distortion due to capacitor non linearity.

For example, 100 µF degraded the THD at 10 Hz from less than 0.0006% to 0.0011%. Band limiting should be done earlier, with non electrolytic capacitors. Protection diode $D_1$ prevents damage to $C_2$ if the amplifier suffers a fault that makes it saturate negatively; it looks unlikely but causes no measurable distortion.[10] $C_7$ provides some stabilising phase advance and limits the closed loop bandwidth; $R_{20}$ prevents it upsetting $Tr_3$.

The VAS stage is enhanced by an emitter follower inside the Miller compensation loop, so that the local NFB which linearises the VAS is increased by augmenting total VAS beta, rather than by increasing the collector impedance by cascoding. The extra local NFB effectively eliminates VAS nonlinearity (Distortion 2).

Increasing VAS beta like this presents a much lower collector impedance than a cascode stage due to the greater local feedback. The improvement appears to make a VAS buffer to eliminate Distortion 4 (loading of VAS collector by the nonlinear input impedance of the output stage) unnecessary. $C_{dom}$ is relatively high at 100 pF, to swamp transistor internal capacitances and circuit strays, and make the design predictable. The slew rate calculates as 40 V/µsec. The VAS collector load is a standard current source, to avoid the uncertainties of bootstrapping.

## Quiescent current stability

Since almost all the THD from a *blameless* amplifier is crossover, keeping the quiescent optimal is essential. Quiescent stability requires the bias generator to cancel out the $V_{be}$ variations of four junctions in series; those of two drivers and two output devices. Bias generator $Tr_{13}$ is the standard $V_{be}$ multiplier, modified to make its voltage more stable against variations in the current through it. These occur because the biasing of $Tr_5$ does not completely reject rail variations: its output current drifts initially due to heating thus changing its $V_{be}$. Keeping a Class B quiescent stable is hard enough at the best of times, and so it makes sense to keep these extra factors out of the equation.

The basic $V_{be}$ multiplier has an incremental resistance of about 20 Ω; in other words its voltage changes by 1 mV for a 50 µA drift in standing current. Adding $R_{14}$ converts this to a gently peaking characteristic that can be made perfectly flat at one chosen current; see Figure 5. Setting $R_{14}$ to 22 Ω makes the voltage peak at 6 mA, and standing current now must deviate from this value by more than 500 µA for a 1 mV bias change.

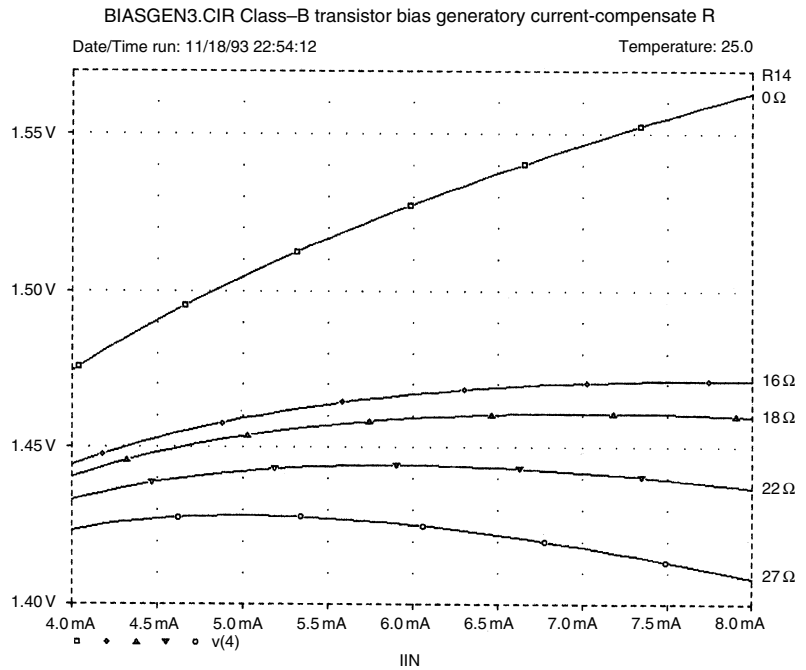



**Figure 5** *Spice plot of the voltage-peaking behaviour of a current-compensated bias generator.*

The $R_{14}$ value needs to be altered if $Tr_5$ is run at a different current. For example, 16 Ω makes the voltage peak at 8 mA instead. If TO3 outputs are used, the bias generator should be in contact with the top or can of one of the output devices, rather than the heatsink, as this is the fastest and least attenuated source for thermal feedback.

## Output stage

The output stage is a standard double emitter follower apart from the connection of $R_{15}$ between the driver emitters without connection to the output rail. This gives quicker and cleaner switch-off of the outputs at high frequencies; this may be significant from 10 kHz upwards dependent on transistor type. Speed up capacitor $C_5$ improves the switch-off action. $C_6$, $R_{18}$ form the Zobel network while $L_1$, damped by $R_{19}$, isolates the amplifier from load capacitance.

Figure 6 shows the 50 W/8 Ω distortion performance, about 0.001% at 1 kHz, and 0.006% at 10 kHz. The measurement bandwidth makes a big difference to the appearance, because what little distortion is present is

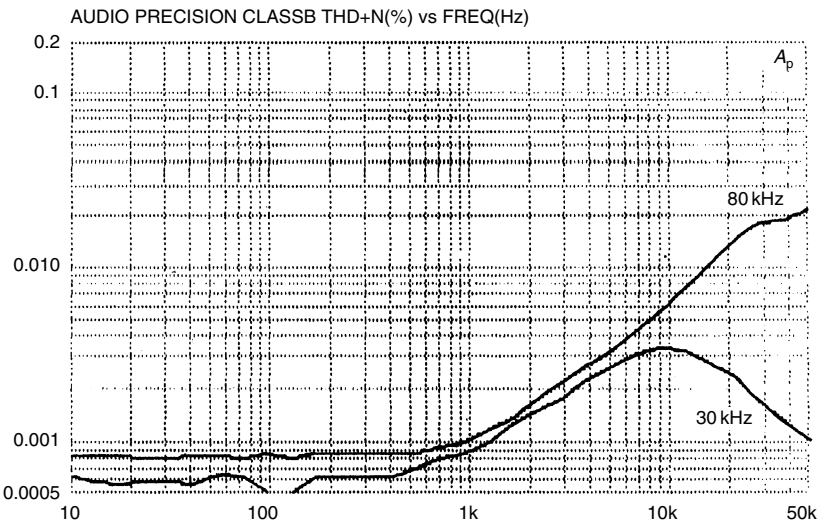



**Figure 6** *Class B amplifier: THD performance at* 50 W/8-ohm*; measurement bandwidths* 30 kHz *and* 80 kHz.

crossover derived, and so high order. It rises at 6 dB/octave, the rate at which feedback factor falls. The crossover glitches emerge from the noise, like Grendel from the marsh, as the test frequency increases above 1 kHz. There is no precipitous THD rise in the ultrasonic region, and so I suppose this might be called a high speed amplifier.

Note that the zigzags on the LF end of the plot are measurement artifacts, apparently caused by the Audio Precision system trying to winkle distortion from visually pure white noise. Below 700 Hz the residual was pure noise with a level equivalent to approx 0.0006% at 30 kHz bandwidth. The actual THD here must be microscopic.

This performance can only be obtained if all seven of the distortion mechanisms are properly addressed; Distortions 1–4 are determined by the circuit design, but the remaining three depend critically on physical layout and grounding topology.

Figure 7 shows the startling results of applying two pole compensation to the amplifier. $C_3$ remains 100 pF, while $C_{p2}$ was 220 pF and $R_p$ 1 kΩ. The extra NFB does its work extremely well, the 10 kHz THD dropping to 0.0015%, while the 1 kHz figure can only be guessed at. There were no unusual signs of instability on the bench, but I have not tried a wide range of loads.

This experimental amplifier was rebuilt with three alternative output stages: the simple quasi-complementary, the quasi-Baxandall and the CFP. The results for both single and two pole compensation are shown in Figures 8, 9, and 10. The simple quasi-complementary generates more

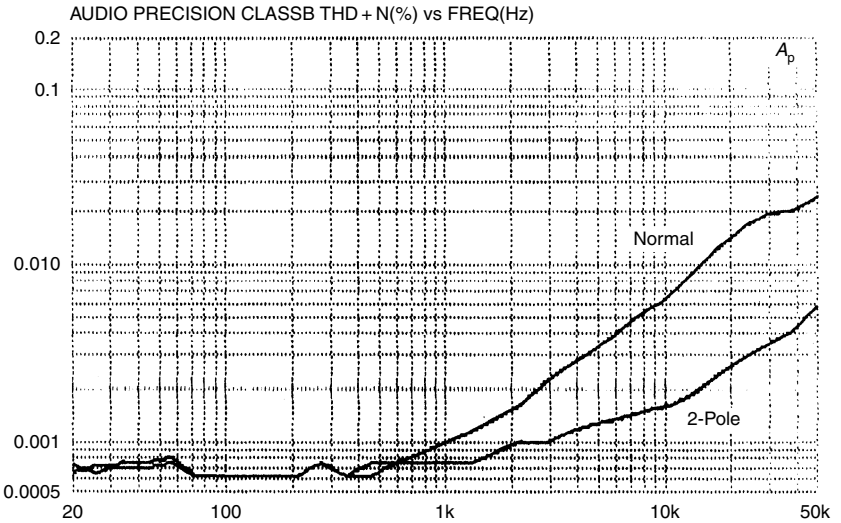



**Figure 7** *The dramatic THD improvement obtained by converting the Class B amplifier to two pole.*

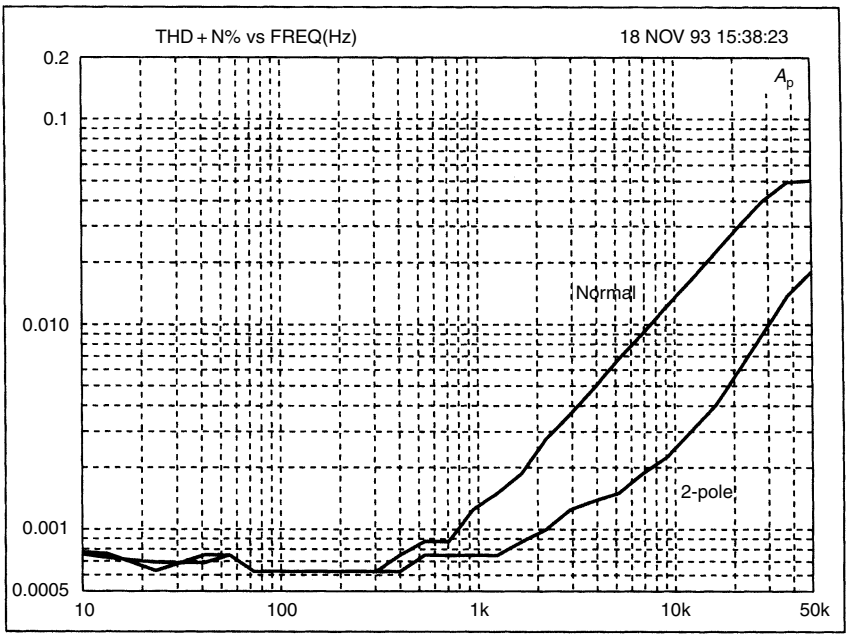



**Figure 8** *Class B amplifier with simple quasi-complementary output. Lower trace is for two pole compensation (*80 kHz *bandwidth).*

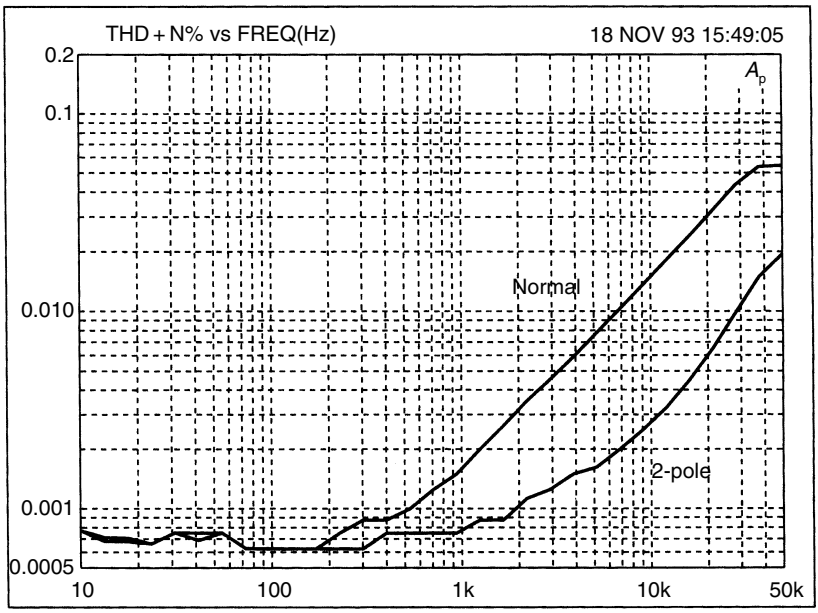



**Figure 9** *Class B amplifier with quasi-complementary plus Baxandall diode output. Lower trace is the two pole case (*80 kHz *bandwidth).*

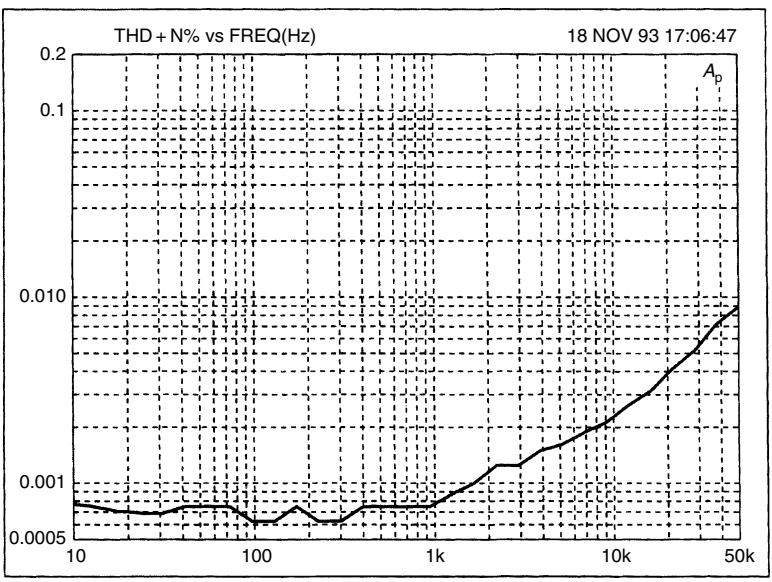



**Figure 10** *Class B amplifier with complementary feedback pair (CFP) output stage.*

crossover distortion, as expected, and the quasi-Baxandall version is not a lot better, due to remaining asymmetry around the crossover region. The CFP gives even lower distortion than the original EF-II output. Figure 10 shows only the result for single pole compensation; in this case the improvement with two pole was marginal and the trace is omitted for clarity.

*The AP plots in earlier parts of this series were mostly done with an amplifier similar to Figure 4, though of higher power. Main differences were the use of a cascode-VAS with a buffer, and a CFP output to minimise distracting quiescent variations. Measurements at powers above* 100 W/8 Ω *used a version with two paralleled output devices.*

## References


1. Otala, 'An audio power amplifier for ultimate quality requirements', *IEEE Trans on Audio & Electroacoustics*, Vol AU-21, No. 6, December 1973.
2. Baxandall, 'Audio power amplifier design. Part 4', *Wireless World*, July 1978, p 76.
3. Takahashi *et al.* 'Design & construction of high slew-rate amplifiers', *AES* 60th Convention, Preprint No. 1348 (A-4) 1978.
4. Self, 'Crossover Distortion & Compensation', Letters, *Electronics & Wireless World*, August 1992, p 657.
5. Widlar, 'A monolithic power op-amp', *IEEE Journal of Solid-State Circuits*, Vol 23, No 2, April 1988.
6. Bonello, 'Advanced negative feedback design for high performance amplifiers', *AES* 67th Convention, Preprint No. 1706 (D-5) 1980.
7. Pernici, 'A CMOS low-distortion amplifier with double-nested et al miller compensation', *IEEE Journal of Solid-State Circuits*, July 1993, p 758.
8. 'Fast compensation extends power bandwidth', Linear Brief 4, National Semiconductor Linear Applications Handbook 1991.
9. Feucht, *Handbook of Analog Circuit Design*, Academic Press 1990, p 264.
10. Self, *'An Advanced Preamplifier', Wireless World*, November 76, p 43.

# 23 Distortion in power amplifiers, Part VIII: Class A amplifiers

*March 1994*

There are two salient facts about Class A amplifiers: they are inefficient; they give the best possible distortion performance. The quiescent dissipation of the classic Class A amplifier is equal to twice the maximum output power, making massive output power impractical. But the nature of our hearing means that the power of an amplifier must be considerably increased to sound significantly louder. It is well known that power in watts must be quadrupled to double sound pressure level (SPL), but this is *not* the same as doubling subjective loudness; this is measured in Sones rather than dB above threshold, and some researchers have reported that doubling subjective loudness requires a 10 dB rather than 6 dB rise in SPL, implying that amplifier power must be increased tenfold, rather than merely quadrupled.[1] This may help to put worries about amplifier size into perspective . . .

There is an attractive simplicity about class A. Most of the distortion mechanisms studied so far stem from class B, and we can thankfully forget crossover and switchoff phenomena (Distortions 3b, 3c), non-linear VAS loading, (Distortion 4) injection of supply-rail signals, (Distortion 5) induction from supply currents, (Distortion 6), and erroneous feedback connections. (Distortion 7) Beta-mismatch in the output devices can also be ignored.

## The art of compromise

The only real disadvantage of class A is inefficiency, so inevitably efforts have been made to compromise between A and B. As compromises go, traditional class AB is not a happy one because, when the AB region is entered, the step change in gain generates significantly greater high order distortion than that from optimally biased class B. However, a well-designed

AB amplifier will give pure class A performance below the AB threshold, something a class B amp cannot do.

Another compromise is the so-called non-switching amplifier, with its output devices clamped to pass a minimum current. However, it is not immediately obvious that a sudden halt in current-change as opposed to complete turn-off makes for a better crossover region. Those residual oscillograms that have been published seem to show that some kind of discontinuity still exists at crossover.[2]

A potential problem is the presence of maximum ripple on the supply rails at zero signal output; the PSRR must be taken seriously if good noise and ripple figures are to be obtained. This problem can be simply solved by the measures proposed for class B designs.

There is a kind of canonical sequence of efficiency improvement in class A amplifiers. The simplest kind is single-ended and resistively loaded, as at Figure 1(a). When it sinks output current, there is an inevitable voltage drop across the emitter resistance, limiting the negative output capability, and resulting in an efficiency of 12.5% (erroneously quoted in at least one textbook as 25%, apparently on the grounds that power not dissipated in silicon doesn't count) This would be of purely theoretical interest – and not much of that – except that a single ended design has recently appeared. This reportedly produces a 10 W output for a dissipation of 120 W, with output swing predictably curtailed in one direction.[3]

A better method – constant current class A – is shown in Figure 1(b). The current sunk by the lower constant current source is no longer related to the voltage across it, and so the output voltage can approach the negative rail with a practicable quiescent current. (Hereafter shortened to '$I_q$') Maximum efficiency is doubled to 25% at maximum output; for an example with 20 W output (and a big fan) see Ref. 4. Some versions (Krell) make the current source value switchable, controlling it with a kind of noise gate.

Push-pull operation once more doubles full-power efficiency, producing a more practical 50%; most commercial class A amplifiers have been of this type. Both output halves now swing from zero to twice the $I_q$, and least voltage corresponds with maximum current, reducing dissipation. There is also the intriguing prospect of cancelling the even-order harmonics generated by the output devices.

There are several ways to induce push-pull action. Figures 1(c), (d) show the lower constant current source replaced by a voltage controlled current source. This can be driven directly by the amplifier forward path, as in Figure 1(c),[5] or by a current control negative feedback loop, as at Figure 1(d).[6] The first of these methods has the drawback that the stage generates gain, phase splitter $Tr_1$ doubling as the VAS; hence there is no circuit node that can be treated as the input to a unity gain output stage, making the circuit hard to analyse, as VAS distortion cannot be separated

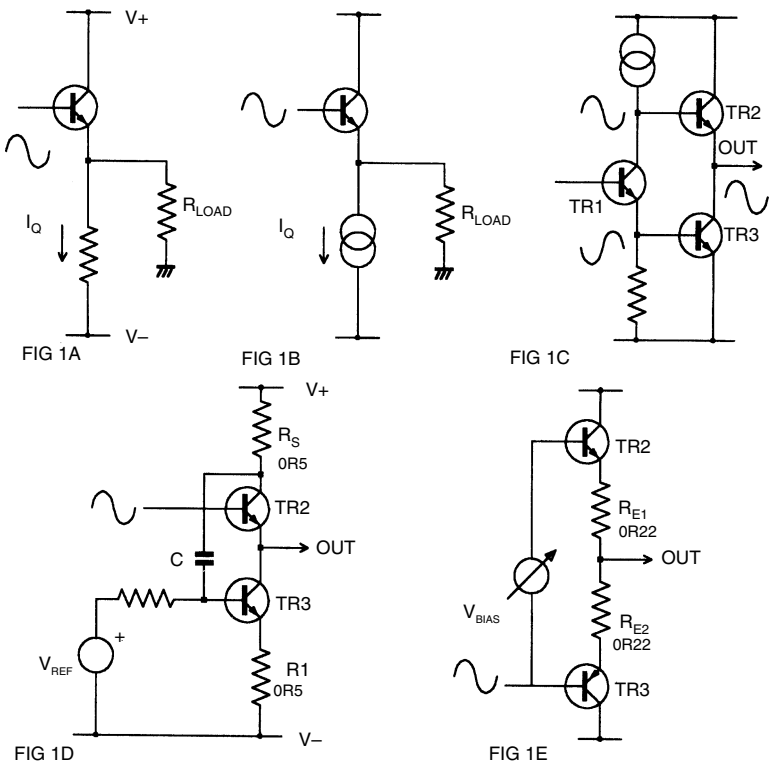



**Figure 1** *The major class A configurations. 1c, 1d and 1e are push-pull variants, 1e being simply a class B stage with higher* $V_{bias}$.

from output stage non-linearity. There is also no guarantee that upper and lower output devices will be driven appropriately for class A if the effective quiescent varies by more than 40% over the cycle.[5]

The second push-pull method in 1d is more dependable, and I can vouch that it works well. The disadvantage with the simple form shown is that a regulated supply is required to prevent rail ripple from disrupting the current loop control. Designs of this type have a limited current control range. In Figure 1(d), $Tr_3$ cannot be turned on further once the upper device is fully off – so the voltage controlled current source will not be able to respond to an unforeseen increase in the output loading. If this happens there is no way of resorting to class AB to keep the show going and the amplifier will show some form of asymmetrical hard clipping.

The best push-pull stage seems to be that in Figure 1(e), which probably looks rather familiar. Like all the conventional class B stages examined in Chapter 19, this one will operate effectively in push-pull class A if the

bias voltage is sufficiently increased; the increase over class B is typically 700 mV, dependant on the value of the emitter resistors. For an example of high biased class B see Ref. 7. This topology has the great advantage that, when confronted with an unexpectedly low load impedance, it will operate in class AB. The distortion performance will be inferior not only to class A but also to optimally biased class B, once above the AB transition level, but can still be made very low by proper design.

Although the push-pull concept has a maximum efficiency of 50%, this is only true at maximum sinewave output. Due to the high peak/average ratio of music, the true average efficiency probably does not exceed 10%, even at maximum volume before obvious clipping.

Other possibilities are signal controlled variation of the class A amplifier rail voltages, either by a separate class B amplifier, or a modulated switch mode supply. Both approaches are capable of high power output, but involve extensive extra circuitry, and present sent some daunting design problems.

A class B amplifier has a limited voltage output capability, but can be flexible about load impedances, as more current will be simply turned on when required. However, class A has also a current limitation, after which it enters class AB, and so loses its *raison d'etre*. The choice of quiescent value has a major effect on thermal design and parts cost so a clear idea of load impedance is important. The calculations to determine the required $I_q$ are straightforward, though lengthy if supply ripple, $V_{ce(sat)}$, and $R_e$ losses, etc., are all considered, so I just give the results here. An unregulated supply with 10,000 μF reservoirs is assumed.

A 20 W/8 Ω amplifier will require rails of approx ±24 V and a quiescent of 1.15A. If this is extended to give roughly the same voltage swing into 4 Ω, then the output power becomes 37 W, and to deliver this in class A the quiescent must increase to 2.16A, almost doubling dissipation. If however full voltage swing into 6 Ω will do, (which it will for many reputable speakers) then the quiescent only needs to increase to 1.5A; from here on I assume a quiescent of 1.6A to give a margin of safety.

## The class A output stage

I consider here only the high biased class B topology, because it is probably the most popular approach, effectively solving the problems presented by the others. Figure 2 shows a Spice simulation of the collector currents in the output devices versus output voltage for the emitter follower configuration, and also the sum of these currents. This sum of device currents is, in principle, constant, but need not be so for low THD. The output signal is the difference of device currents and is not inherently related to the sum. However, a large deviation from this constant sum condition means

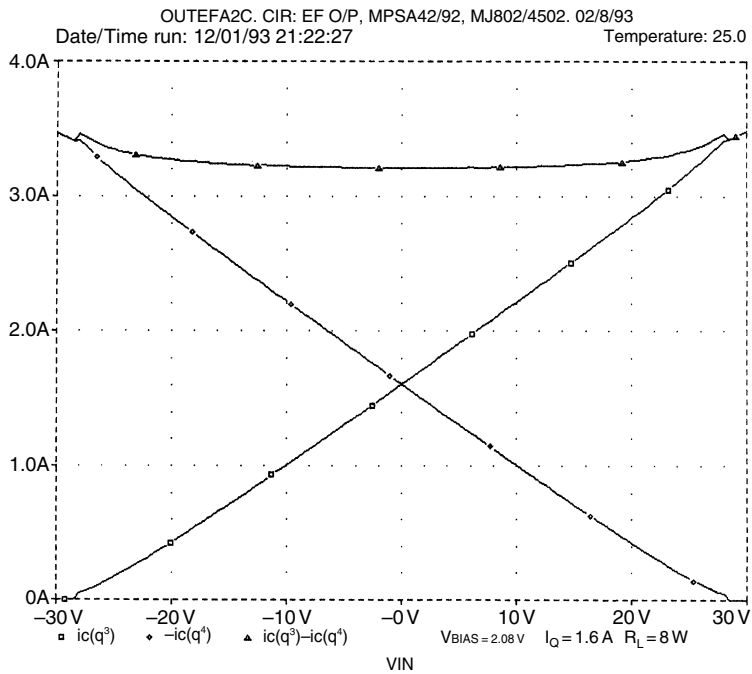



**Figure 2** *How output device current varies in push-pull class A. The sum of the currents is near-constant, simplifying biasing.*

inefficiency, as the stage is conducting more quiescent than it needs to for some part of the cycle. The constancy of this sum is important because it shows that the voltage measured across $R_{e1}$ and $R_{e2}$ together is also effectively constant so long as the amplifier stays in class A. This in turn means that $I_q$ can be simply set with a constant voltage bias generator, in very much the same way as class B.

Figures 3, 4, 5 show Spice gain plots for open loop output stages, with 8 Ω loading and 1.6 A quiescent; the circuitry is exactly as for class B in Part 4. The upper traces show class A gain, and the lower traces gain under optimal class B bias for comparison. Figure 3 shows an emitter follower output, Figure 4(a) simple quasi complementary stage, and Figure 5(a) CFP output.

We would expect class A stages to be more linear than B, and they are Harmonic and THD figures for the three configurations, at 20 V peak, are shown in Table 1. There is absolutely no gain wobble around OV, and push-pull class A genuinely does cancel even order distortion. Class B only does this in the crossover region, in a partial and unsatisfactory way.

It is immediately clear that the emitter follower has more gain variation, and therefore worse linearity, than the CFP, while the quasi complementary

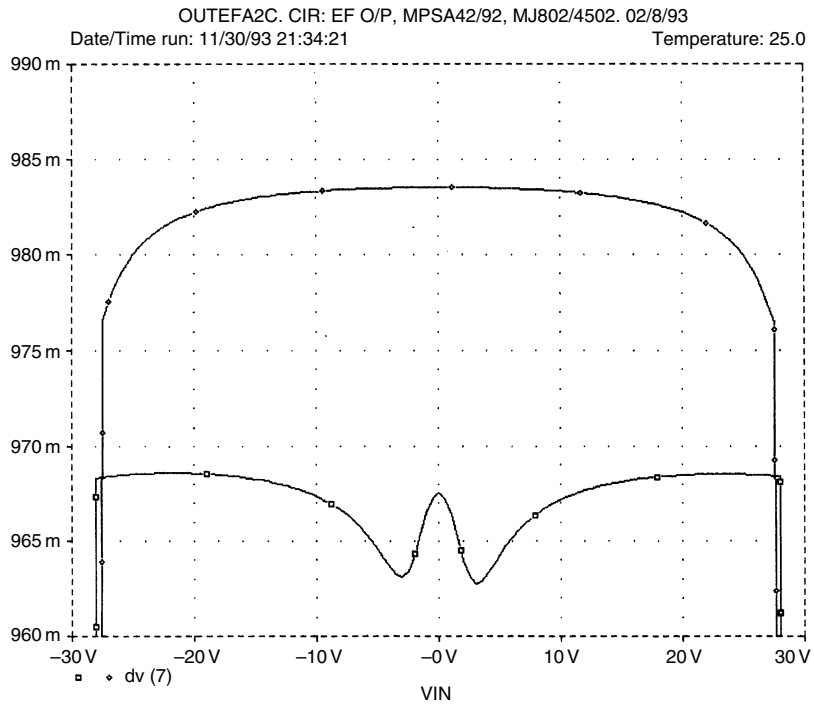



**Figure 3** *Gain linearity of the class A emitter-follower output stage. Load is* 8 Ω*, and quiescent current (Iq) is* 1.6A. *Upper trace class A, lower trace optimal class B.*

circuit shows an interesting mix of the two. The more curved side of the quasi gain plot is on the negative side, where the CFP half of the quasi circuit is passing most of the current. However we know by comparing Figure 3 and Figure 5 that the CFP is the more linear structure. Therefore it appears that the shape of the gain curve is determined by the output half that is turning off, presumably because this shows the biggest $g_m$ changes. The CFP structure maintains $g_m$ better as current decreases, and so gives a flatter gain curve with less rounding of the extremes.

The gain behaviour of these stages is reflected in their harmonic generation; Table 1 reveals that the two symmetrical topologies give mostly odd order harmonics as expected. The asymmetry of the quasi comp version causes a large increase in even order harmonics, and this is reflected in the higher THD figure. Nonetheless the THD figures are still two to three times lower than for their class B equivalents.

If this factor of improvement seems a poor return for the extra dissipation of class A, this is not so. The crucial point about the distortion from a class A output stage is not just that is low, but that it is low order, and

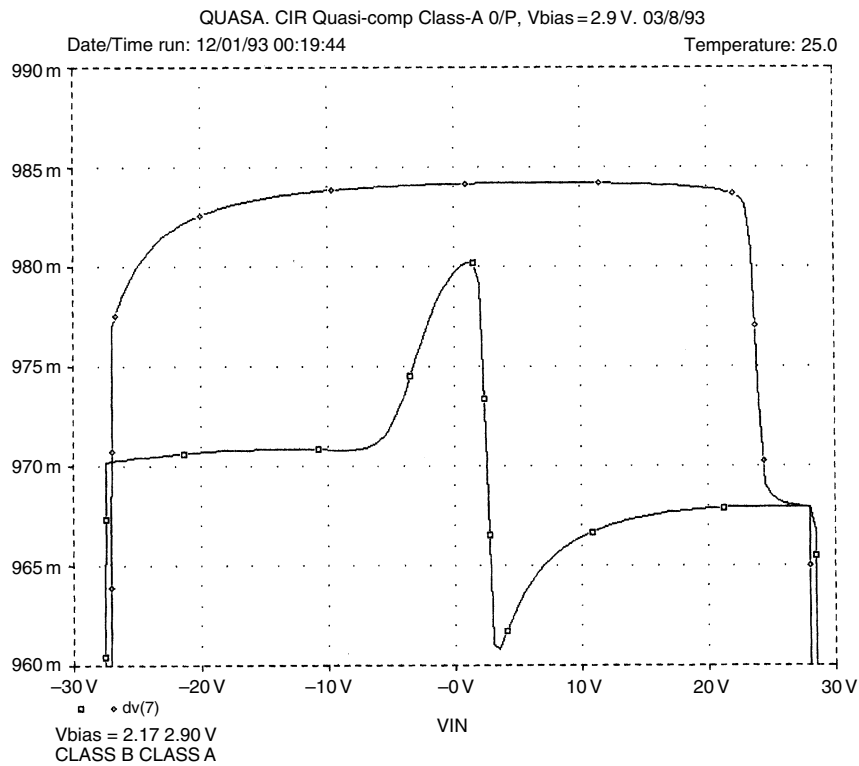



**Figure 4** *Gain linearity of the class A quasi-complementary output stage. Conditions as in Figure 3. Upper trace class A, lower class B.*

so benefits much more from a typical NFB factor that falls with frequency than does high order crossover distortion.

The choice of class A output topology is now simple. For best performance, use the CFP. Apart from greater basic linearity, the effects of output device temperature on $I_q$ are servoed out by local feedback, as in class B. For utmost economy, use the quasi complementary with two NPN devices: these need only a low $V_{ce(max)}$ for a typical class A amp, so here is an opportunity to recoup some of the money spent on heatsinking.

The rules are different from class B; the simple quasi configuration will give first class results with moderate NFB, and adding a Baxandall diode to simulate a complementary emitter follower stage makes little difference to linearity.[7]

It is sometimes assumed that the different mode of operation of class A makes it inherently short circuit proof. This may be true with some configurations, but the high biased type shown here will continue delivering current until it bursts. Overload protection is no less necessary.

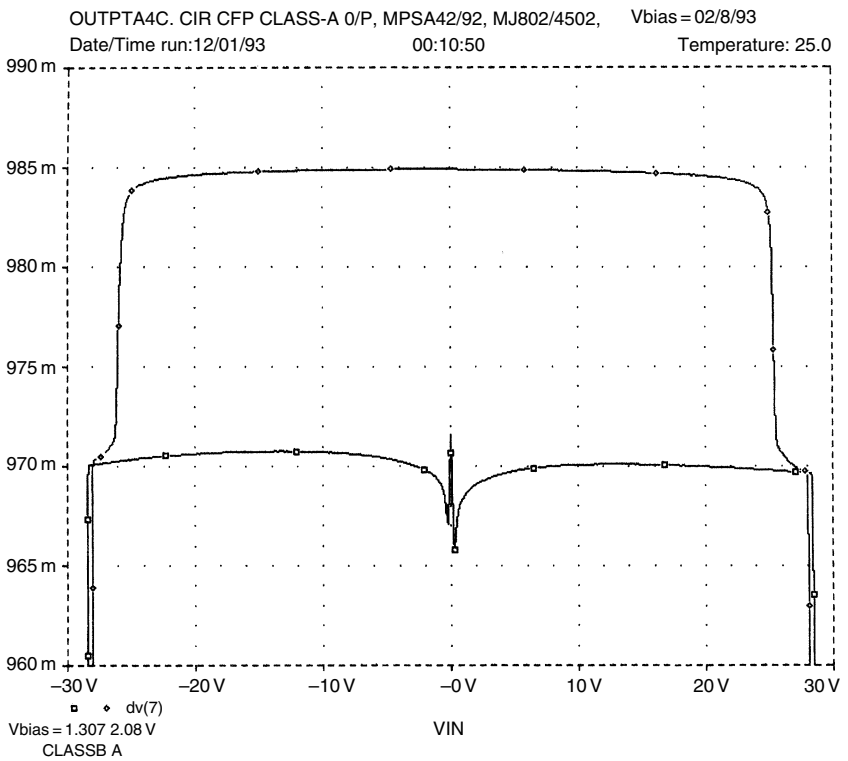



**Figure 5** *Gain linearity of the class A CFP output stage. Upper trace class A, lower trace class B.*

**Table 1**

| *Harmonic* | *Emitter Follower (%)* | *Quasi-Comp (%)* | *CFP Output (%)* |
|---|---|---|---|
| Second | 0.00012 | 0.0118 | 0.00095 |
| Third | 0.0095 | 0.0064 | 0.0025 |
| Fourth | 0.00006 | 0.0011 | 0.00012 |
| Fifth | 0.00080 | 0.00058 | 0.00029 |
| THD | 0.0095 | 0.0135 | 0.0027 |

THD is calculated from the first nine harmonics, though levels above the fifth are very small

## Quiescent control systems

Unlike class B, precise control of quiescent current is not required to optimise distortion. For good linearity there just has to be enough of it. However, $I_q$ must be under some control to prevent thermal runaway,

particularly if the emitter follower output is used, and an ill conceived controller can ruin the THD. There is also the point that a precisely held standing current is considered the mark of a well bred class A amplifier; a quiescent that lurches around like a drunken sailor does not inspire confidence.

Thermal feedback from the output stage to a standard $V_{be}$ multiplier bias generator will work,[8] and should be sufficient to prevent run-away. However, unlike class B, class A gives the opportunity of tightly controlling $I_q$ by negative feedback. This is profoundly ironic because now that we can precisely control $I_q$, it is no longer critical. Nevertheless it seems churlish to ignore the opportunity.

There are two basic methods of feedback current control. In the first, the current in one output device is monitored, either by measuring the voltage across *one* emitter resistor, ($R_s$ in Figure 6(a)), or by a collector sensing resistor. The second method monitors the sum of the device currents, which as described above, is constant in class A.

The first method as implemented in Figure 6(a)[7] compares the $V_{be}$ of $Tr_4$ with the voltage across $R_s$, with filtering by $R_F$, $C_F$. If quiescent is excessive, then $Tr_4$ conducts more, turning on $Tr_5$ and reducing the bias voltage between points A and B.

In Figure 6(b), which uses the voltage controlled current source approach, the voltage across collector sensing resistor $R_s$ is compared with $V_{ref}$ by $Tr_4$, the value of $V_{ref}$ being chosen to allow for $Tr_4 V_{be}$.[9] Filtering is once more by $R_F$, $C_F$.

For either Figure 6(a) or 6(b), the current being monitored contains large amounts of signal, and must be low pass filtered before being used for control purposes. This is awkward as it adds one more time constant to worry about if the amplifier is driven into asymmetrical clipping, and implies the desirability of large electrolytic capacitors to minimise the a.c. voltage drop across the sense resistors. In the case of collector sensing there are unavoidable losses in the extra sense resistor. It is my experience that imperfect filtering can produce a serious rise in distortion.

The better way is to monitor current in *both* emitter resistors. As explained above, the voltage across both is very nearly constant, and in practice filtering is unnecessary. An example of this approach is shown in Figure 6(c), based on a concept originated by Nelson Pass.[10] Here $Tr_4$ compares its own $V_{be}$ with the voltage between X and B; excessive quiescent turns on $Tr_4$ and reduces the bias directly. Diode $D$ is not essential to the concept, but usefully increases the current feedback loop gain; omitting it more than doubles $I_q$ variation with $Tr_7$ temperature in the Pass circuit.

The trouble with this method is that $Tr_3 V_{be}$ directly affects the bias setting, but is outside the current control loop. A multiple of $V_{be}$ is established between X and B, when what we really want to control is the voltage between X and Y. The temperature variations of $Tr_4$ and $Tr_3 V_{be}$ partly

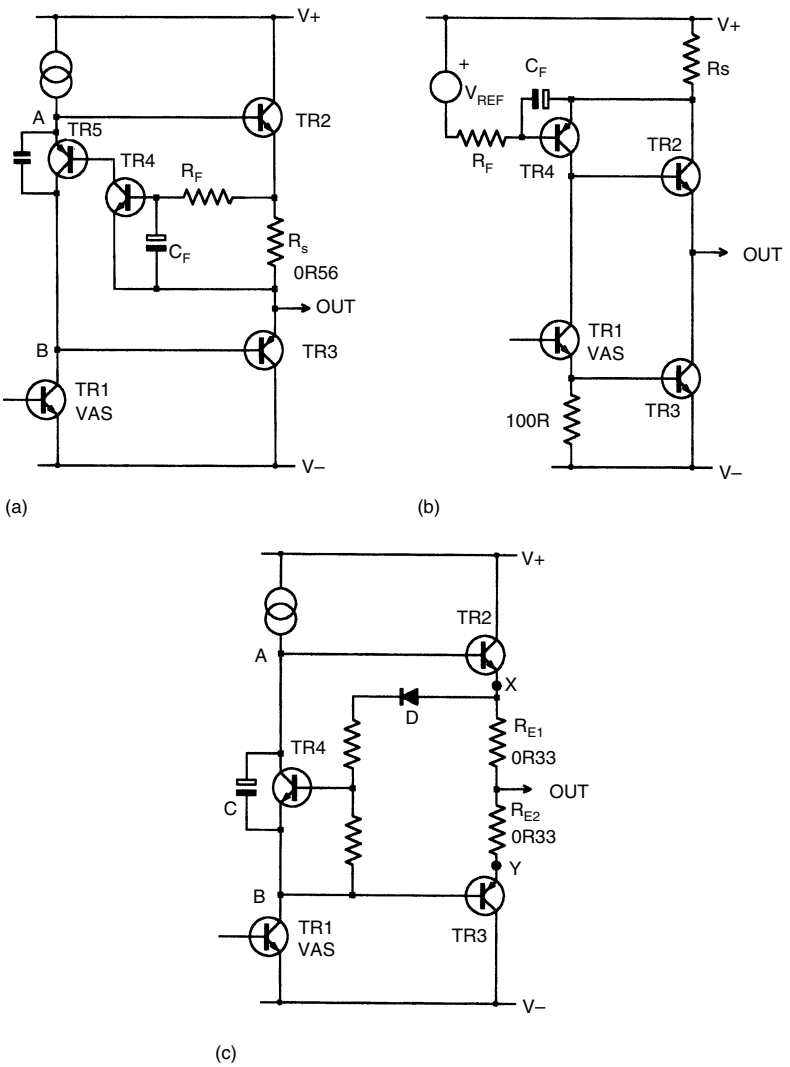



**Figure 6** *Quiescent current-control systems. Only that at Figure 6(c) avoids the need to low pass filter the control signal; C simply provides feed forward to speed up signal transfer to* $\mathrm{Tr}_2$.

cancel, but only partly. This method is best used with a CFP or quasi output so that the difference between Y and B depends only on the driver temperature, which can be kept low. The 'reference' is $Tr_4 V_{be}$, which is itself temperature dependent. Even if it is kept away from the hot bits it will react to ambient temperature changes, and this explains the poor performance of the Pass method for global temperature changes (Table 2).

**Table 2** $I_q$ *change per °C change in temperature*

| | *Changing $Tr_7$ temperature only* | *Changing Global temperature (%)* |
|---|---|---|
| Quasi + $V_{be}$ mult | +0.112% | −0.43 |
| Pass: as Flg. 6c | +0.0257 | −14.1 |
| Pass: no dlode D | +0.0675 | −10.7 |
| New system: | +0.006% | −0.038 |

(assuming 0.22 Ω emitter resistors and 1.6A $I_q$.)

To solve this problem, I would like to introduce the novel control method in Figure 7. We need to compare the floating voltage between X and Y with a fixed reference, which sounds like a requirement for two differential amplifiers. This can be reduced to one by sitting the reference $V_{ref}$ on point Y. This is a very low impedance point and can easily swallow a reference current of 1 mA or so. A simple differential pair $T_{r15,16}$ then compares the reference voltage with that at point X: excess quiescent turns on $Tr_{16}$, causing $Tr_{13}$ to conduct more and reducing the bias voltage.

The circuitry looks enigmatic because of the high impedance of $Tr_{13}$ collector would seem to prevent signal from reaching the upper half of the output stage; this is in essence true, but the vital point is that $Tr_{13}$ is part of a NFB loop that establishes a voltage at A that will keep the bias voltage between A and B constant. This comes to the same thing as maintaining a constant $V_{bias}$ across $Tr_{13}$. As might be imagined, this loop does not shine at transferring signals quickly, and this duty is done by feedforward capacitor $C_4$.

Without it, the loop (rather surprisingly) works correctly, but HF oscillation at some part of the cycle is almost certain. With $C_4$ in place the current loop does not need to move quickly, since it is not required to transfer signal but rather to maintain a DC level.

The experimental study of $I_q$ stability is not easy because of the inaccessibility of junction temperatures. Professional Spice implementations like *PSpice* allow both the global circuit temperature and the temperature of individual devices to be manipulated; this is another aspect where simulators shine. The exact relationships of component temperatures in an amplifier is hard to predict: I show here just the results of changing the global temperature of all devices, and changing the junction temp of $Tr_7$ alone (Figure 7) with different current controllers. $Tr_7$ will be one of the hottest transistors and unlike $Tr_9$ it is not in a local NFB loop, which would greatly reduce its thermal effects.

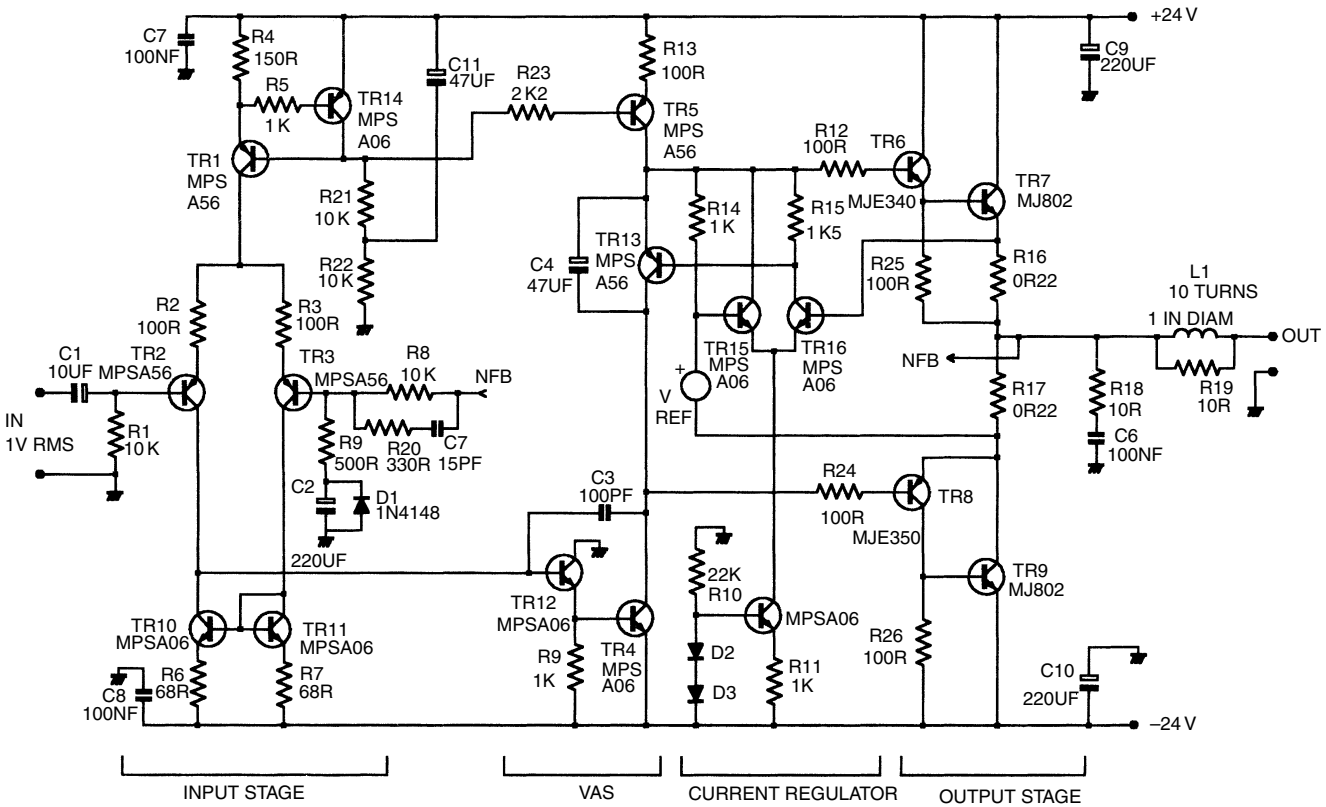



**Figure 7** *A Blameless* 20 W *class A power amplifier, using the novel current-control system.*

## A new class A design

The full circuit diagram shows a 'blameless' 20 W/8 Ω class A power amplifier. This is as close as possible in operating parameters to the previous class B design to aid comparison. In particular the NFB factor remains 30 dB at 20 kHz. The front end is as for the class B version, which should not be surprising as it does exactly same job, input Distortion 1 being unaffected by output topology.

As before the input pair uses a high tail current, so that $R_{2,3}$ can be introduced to linearise the transfer characteristic and set the transconductance. Distortion 2 (VAS) is dealt with as before, the beta enhancer $Tr_{12}$ increasing the local feedback through $C_{dom}$. There is no need to worry about Distortion 4 (non-linear loading by output stage) as the input impedance of a class A output, while not constant, does not have the sharp variations shown by class B.

The circuit uses a standard quasi output. This may be replaced by a CFP stage without problems. In both cases the distortion is extremely low but, gratifyingly, the CFP proves even better than the quasi, confirming the simulation results for output stages in isolation.

The operation of the current regulator $Tr_{13,15,16}$ has already-been described. Using a band gap reference, it holds a 1.6 A $I_q$ to with in ±2 mA from a second or two after switch on. Looking at Table 2, there seems no doubt that the new system is effective.

As before an unregulated power supply with 10,000 μF reservoirs was used, and despite the higher prevailing ripple, no PSRR difficulties were encountered once the usual decoupling precautions were taken.

## Performance

The closed loop distortion performance (with conventional compensation) is shown in Figure 8 for the quasi comp output stage, and in Figure 9 for a CFP output version. The THD residual is pure noise for almost all of the audio spectrum, and only above 10 kHz do small amounts of third harmonic appear. The suspected source is the input pair, but this so far remains unconfirmed.

The distortion generated by the class B and A design examples is summarised in Table 3, which shows a pleasing reduction as various measures are taken to deal with it. As a final tweak, two pole compensation was applied to the most linear (CFP) of the class A versions, reducing distortion to 0.0012% at 20 kHz, at some cost in slew rate (Figure 10). While this may not be the fabled straight wire with gain, it must be a near relation . . .

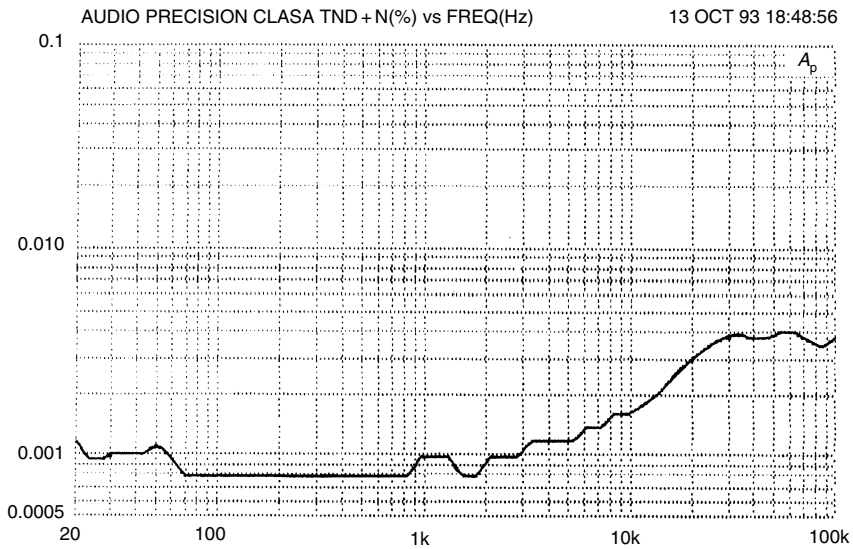



**Figure 8** *Class A amplifier THD performance with quasi-comp output stage. The steps in the LF portion of the trace are measurement artifacts.*

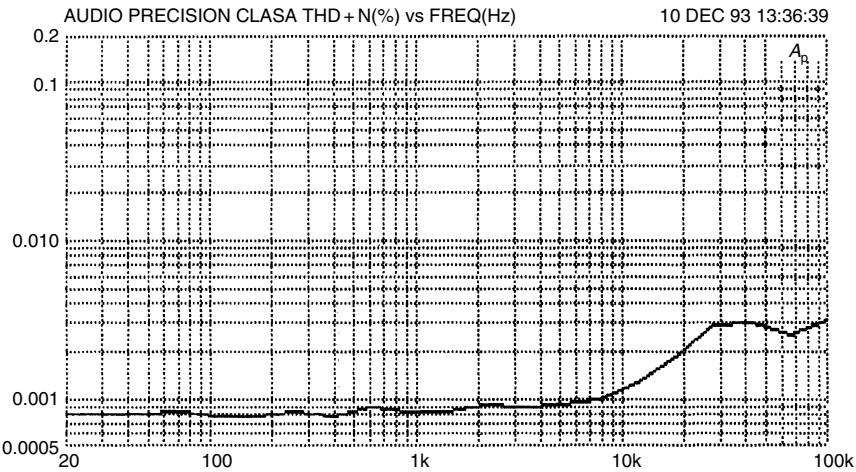



**Figure 9** *Class A distortion performance with CFP output stage.*

## And finally

The techniques in this series have a relevance beyond power amplifiers. Applications obviously include discrete op-amp based preamplifiers[11] and extend to any amplifier offering static or dynamic precision. My philosophy

**Table 3**

| | 1 kHz | 10 kHz | 20 kHz | *Power* |
|---|---|---|---|---|
| class B EF | <.0006% | .0060% | .012% | 50 W |
| class B CFP | <.0006% | .0022% | .0040% | 50 W |
| class B EF 2-pole | <.0006% | .0015% | .0026% | 50 W |
| class A quasl | <.0006% | .0017% | .0030% | 50 W |
| class A CFP | <.0006% | .0010% | .0018% | 20 W |
| class A CFP 2-pole | <.0006% | .0010% | .0012% | 20 W |

(All for 8 Ω loads and 80 kHz bandwidth. Single pole compensation unless otherwise stated.)

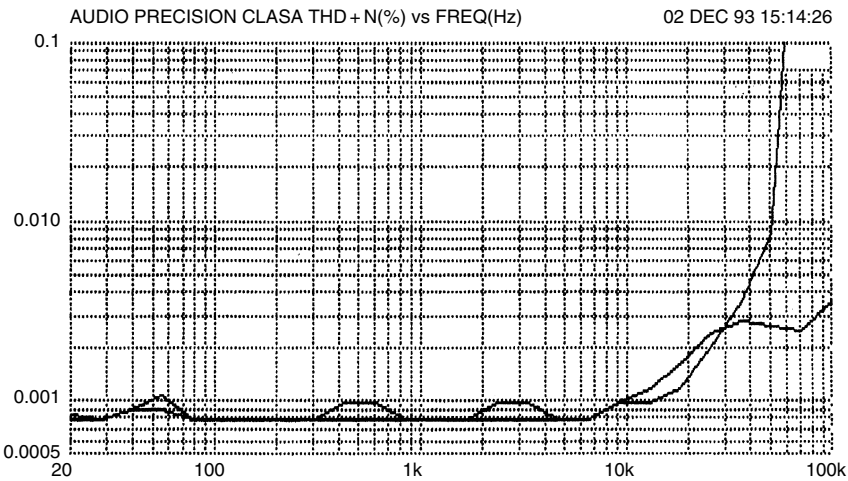



**Figure 10** *Distortion performance for CFP output stage with 2-pole compensation. The THD drops to 0.0012% at* 20 kHz*, but the extra VAS loading has compromised the positive-going slew capability, The 2-pole trace is shown moving off the graph at* 50 kHz.

is that all distortion is bad, and high order distortion is worse . . . $n^2/4$ worse, according to many authorities[12] Digital audio routinely delivers the signal with less than 0.002% THD, and I can earnestly vouch for the fact that analogue console designers work hard to keep the distortion in long and complex signal paths down to similar levels. I think it an insult to allow the very last piece of electronics in the chain to make nonsense of these efforts.

I do not believe that an amplifier yielding 0.001% THD is going to sound much better than another generating 0.002%. However, if there is ever a doubt as to what level of distortion is perceptible, then using the

techniques I have presented, it should be possible to reduce the THD below the level at which there can be any rational argument.

I am painfully aware of the school of thought that regards low distortion as inherently immoral, but this is to confuse electronics with religion. The implication is that very low THD can only be obtained by huge global NFB factors which in turn require heavy dominant pole compensation that severely degrades slew rate. The obvious flaw in this argument is that, once the compensation is applied, the amplifier no longer has a large global NFB factor: Its distortion performance presumably reverts to mediocrity, further burdened with a slew rate of 4 V per fortnight.

To me low distortion has its own aesthetic appeal. All of the linearity enhancing strategies examined in this series are of minimal incremental cost to existing designs with the possible exception of using class A. There seems to be no reason to not use them.

## References

1. Moore, B.J. *An Introduction To The Psychology of Hearing*, Academic Press, 1982, pp 48–50.
2. Tanaka, S. 'A new biasing circuit for class B operation', JAES January/February 1981, p 27.
3. Fuller, S. Private communication.
4. Nelson Pass 'Build a class A amplifier', *Audio*, February 1977, p 28 (constant current).
5. Linsley Hood, J. 'Simple class A Amplifier', *Wireless World*, April 1969, p 148.
6. Self, D. 'High-performance preamplifier', *Wireless World*, February 1979, p 41.
7. Nelson-Jones, L. 'Ultra-low distortion class A amplifier', *Wireless World*, March 1970, p 98.
8. Giffard, T. 'Class A power amplifier', *Elektor*, November 1991, p 37.
9. Linsley-Hood, J. 'High-quality headphone amp', *HiFi News & RR*, January 1979, p 81.
10. Nelson Pass. 'The pass/A40 power amplifier', *The Audio Amateur*, #4, 1978, p 4 (Push-pull).
11. Self, D. 'Advanced Preamplifier Design', *Wireless World*, November 1976, p 41.
12. Moir, J. 'Just detectable distortion levels', *Wireless World*, February 1981, p 34.

# 24 Power amplifier input currents and their troubles

*May 2003*

This article grew out of some work that I was doing for a well-known amplifier company. I had produced a nice low-distortion design, which was to a great extent a straightforward application of the Blameless amplifier design methodology described in the *Distortion In Power Amplifier* series. However, a late change to the specification of the product – a thing not wholly unknown in the world of audio engineering – meant that a resistive network had to be added immediately before the power amplifier stage. The effective source resistance of the network was, if memory serves, 2 k$\Omega$, and when you have perused the following article you will understand that the effects on both the hum and distortion performance were most unwelcome.

Adding a 5532 buffer stage between resistive network and amplifier would have been a quick fix, but running another op-amp from the very limited amount of $\pm$15 V power available was going to be awkward, and PCB area in the right place was also a very scarce resource. The 5532 is a low-noise op-amp, but it is not as quiet as the pair of discrete transistors in the power amplifier input section, and the overall noise performance would definitely have suffered.

It was therefore time to look a little more closely at the exact mechanisms by which the source resistance was causing trouble, in the hope that more elegant ways of retrieving the original performance could be found. They were, and this chapter tells the story of how those mechanisms were uncovered, and rendered less troublesome.

When power amplifiers are measured, the input is normally driven from a low impedance signal generator. Some testgear, such as the much-loved Audio Precision System-1, has selectable output impedance options of 50, 150, and 600 Ω. The lowest value available is almost invariably used because (1) it minimises the Johnson noise from the source resistance; (2) it minimises level changes due to loading by the amplifier input impedance.

This is all very sensible, and exactly the way I do it myself – 99% of the time. There is however two subtle effects that can be missed if the amplifier is always tested this way. These are: distortion caused by the non-linear input currents drawn by the typical power amplifier, and hum caused by ripple modulation of the same input currents.

Note that this is not the same effect as the excess distortion produced by FET-input opamps when driven from significant source impedances; this is due to their non-linear input capacitances to the IC substrate, and has no equivalent in power amplifiers made of discrete transistors.

Figure 1 shows both the effects. The amplifier under test was a conventional Blameless design with an EF output stage comprising a single pair of sustained-beta bipolar power transistors; see Figure 2 for the basic circuit. Output power was 50 W into 8 Ω. The bottom trace is the distortion + noise with the usual source impedance of 50 Ω, and the top one shows how much worse the THD is with a source impedance of 3.9 K. Intermediate traces are for 2.2 K and 1.1 K sources. The THD residual shows both second

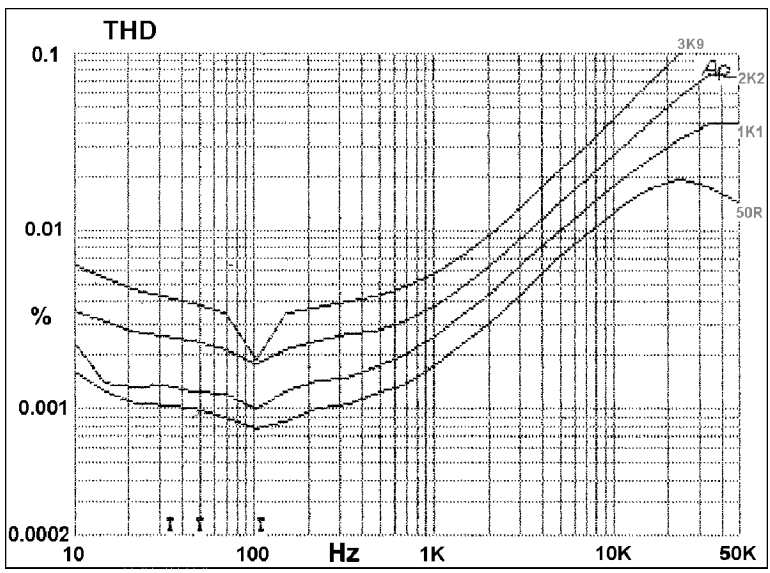



**Figure 1** *Second-harmonic distortion and* 100 Hz *ripple get worse as the source impedance rises from* 50 Ω *to* 3.9 K. 50 W *into* 8 Ω.

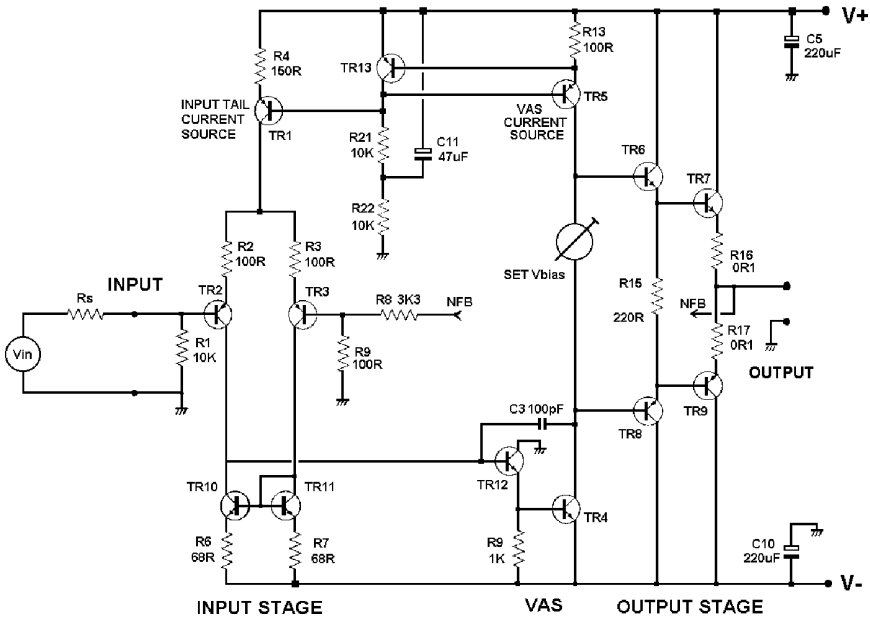



**Figure 2** *Simplified circuit of a typical Blameless power amplifier, with negative-feedback control of VAS current source TR5 by TR13. The bias voltage generated is also used by the input tail source TR1.*

harmonic distortion and 100 Hz ripple components, the latter dominating at low frequencies, while at higher ones the reverse is true. The presence of ripple is signalled by the dip in the top trace at 100 Hz, where distortion products and ripple have partially cancelled. The amount of degradation is proportional to the source impedance.

This is not a problem in most cases, where the preamplifier is driven by an active preamplifier, or by a buffer internal to the power amplifier. Competent preamplifiers have a low output impedance, often around 50–100 Ω, to minimise high-frequency losses in cable capacitance. (I have just been hearing of a system with 10 m of cable between preamp and power amp.)

However, there are two scenarios where the input source resistance is higher than this. If a so-called 'passive preamp' is used then the output impedance is both higher and volume-setting dependent. A 10 K volume potentiometer has a maximum output impedance of one-quarter the track resistance, i.e. 2.5 K, at its mid-point setting. It is also possible for significant source resistance to exist inside the power amplifier for example, there might be an balanced input amplifier, which while it has a very low output impedance itself, may have a resistive gain control network between it and the power amp.

So – we have a problem, or rather two of them. It seems very likely that the input transistor base currents are to blame for both, hence an

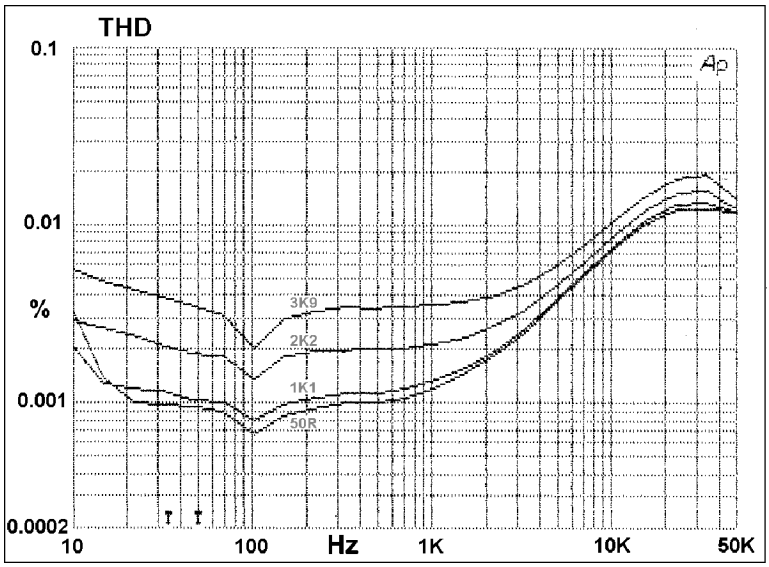



**Figure 3** *There is less introduction of ripple and distortion with high-beta input transistors and the same set of source resistances as Figure 1.*

obvious option is to minimise these currents by using transistors with the highest available beta in the input pair. In this amplifier the input pair were originally ZTX753, with a beta range of 70–200. Replacing these with BC556B input devices (beta range 180–460) gives Figure 3 which shows a useful improvement in THD above 1 kHz; distortion at 10 kHz drops from 0.04% to 0.01%. Our theory that the base currents are to blame is clearly correct. The bottom trace is the reference 50 Ω source plot with the original ZTX753s, and this demonstrates that the problem has been reduced but certainly not eliminated.

The amplifier here is very linear with a low source impedance, and it might well be questioned as to why the input currents drawn are distorted if the output is beautifully distortion-free. The reason is of course that global negative feedback constrains the output to be linear because this is where the NFB is taken from but the internal signals of the amplifier are whatever is required to keep the output linear. The VAS is known to be non-linear, so if the output is sinusoidal the collector currents of the input pair clearly are not. Even if they were, the beta of the input transistors is not constant so the base currents drawn by them would still be non-linear.

It is also possible to get a reduction in hum and distortion by reducing the input pair tail current, but this very important parameter also affects input stage linearity and the slew-rate of the whole amplifier. Figure 4 shows the result. The problem is reduced – though far from eliminated – but the

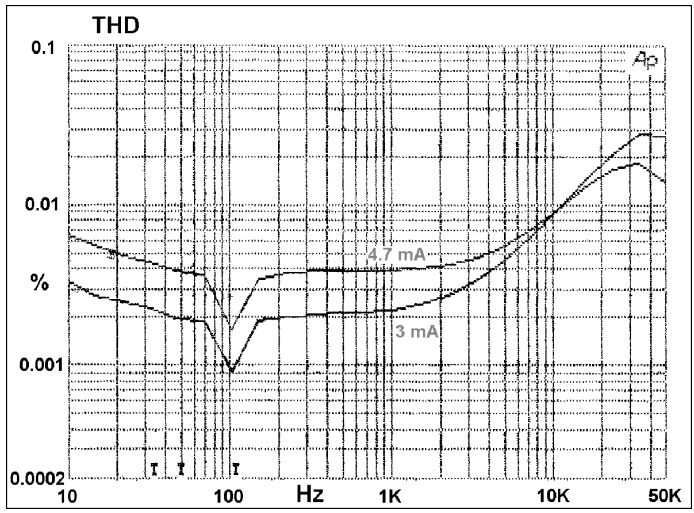



**Figure 4** *Reducing the tail current improves things at low frequencies but increases HF distortion above* 10 kHz. *The notches at* 100 Hz *indicate that the ripple content is still substantial.*

high-frequency THD has actually got worse because of poorer linearity in the input stage. This is not a promising route to follow.

Both ripple and THD effects consequent on the base currents drawn could be eliminated by using FETs instead of bipolars in the input stage. The drawbacks are:

1 Poor Vgs matching, which means that a d.c. servo becomes essential to control the amplifier output d.c. offset. Dual FETs do exist but they are discouragingly expensive.
2 Low transconductance, which means the stage cannot be linearised by local feedback as the raw gain is just not available.
3 Although there is no d.c. gate current, there might well be problems with non-linear input capacitance, as there are with FET-input op-amps.

Once again, not a promising route.

The distortion problem looks rather intractable; one possible total cure is to put a unity-gain buffer between input and amplifier. The snag (for those seeking the highest possible performance) is that any opamp will compromise the noise and distortion of a Blameless amplifier. It is quite correct to argue that this doesn't matter, as any preamp hooked up to the power amp will have opamps in it anyway, but the preamp is a different box, a different project, and possibly has a different designer, so philosophically this does not appeal to everyone. If a balanced input is required then

an opamp stage is mandatory. (unless you prefer transformers, which of course have their own problems.)

The best choice for the opamp is either the commonplace but extremely capable 5532 (which is pretty much distortion-free, but not alas noise-free, though it is very quiet) or the rather expensive but very quiet AD797.

The ripple problem, however, has a more elegant solution. If there is ripple in the input base current, then clearly there is some ripple in the tail current. This is not normally detectable because the balanced nature of the input stage cancels it out. A significant input source impedance upsets this balance, and the ripple appears.

The tail is fed from constant-current source TR1, and this is clearly not a mathematically perfect circuit element. Investigation showed that the cause of the tail-current ripple contamination is Early effect in this transistor, which is effectively fed with a constant bias voltage A tapped off from the VAS negative-feedback current source. (Early effect is the modulation of transistor collector current caused by changing the Vce; as a relatively minor aspect of bipolar transistor behaviour it is modelled by SPICE simulators in a rather simplistic way.) Note that this kind of negative-feedback current-source could control the tail current instead of the VAS current, which might well reduce the ripple problem, but is arranged this way as it gives better positive slewing. Another option is two separate negative-feedback current-sources.

The root cause of our hum problem is therefore the modulation of the Vce of TR1 by ripple on the positive rail, and this variation is easily eliminated by cascoding, as shown in Figure 5. This forces TR1 emitter and collector to move up and down together, preventing Vce variations. It completely eradicates the ripple components, but leaves the input-current distortion unaltered, giving the results in Figure 6, where the upper trace is degraded only by the extra distortion introduced by a 2 K source impedance; the 100 Hz cancellation notch has also disappeared. The reference 50 Ω source plot is below it.

The voltage at A that determines the Vce of TR1 is not critical. It must be sufficiently below the positive supply rail for TR1 to have enough Vce to conduct properly, and it must be sufficiently above ground to give the input pair enough common-mode range. I usually split the biasing chain R21, R22 in half, as shown, so C11 can be used to filter out rail noise and ripple, and biasing the cascode transistor from the mid-point works very well.

It may have occurred to the reader that simply balancing the impedances seen by the two inputs will cancel out the unwanted noise and distortion. This is not very practical as with discrete transistors there is no guarantee that the two input devices will have the same beta. (I know there are such things as dual bipolars, but once more the cost is depressing.) This also implies that the feedback network will have to have its impedance raised

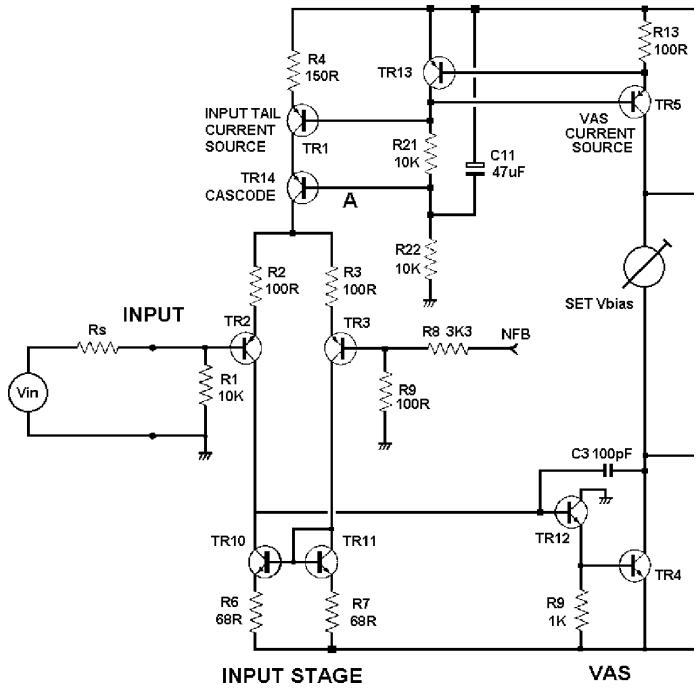



**Figure 5** *Cascoding the input tail; one method of biasing the cascode.*

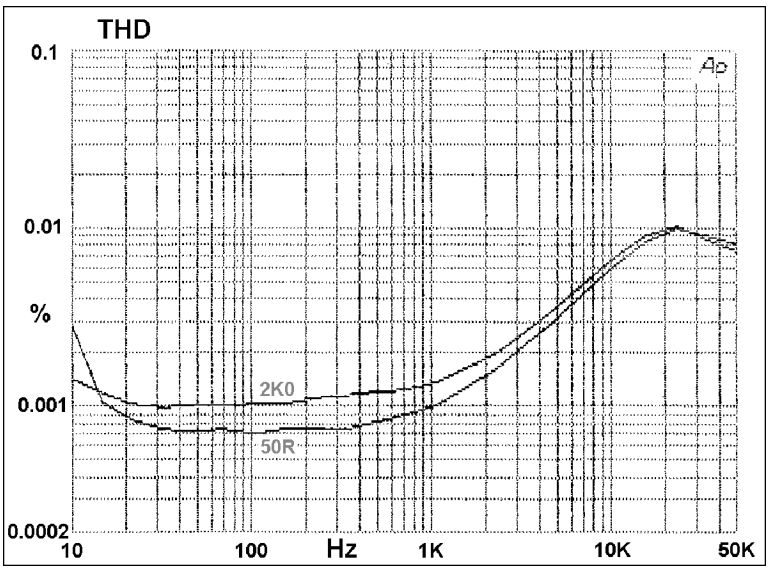



**Figure 6** *Cascoding the input tail removes the ripple problem, but not the extra distortion.*

to equal that at the input, which would give unnecessarily high levels of Johnson noise.

## Conclusions

If the system design requires an opamp at the input, then both hum and distortion problems are removed with no further effort. If not, perhaps because the amplifier must be as quiet as possible, then cascoding the input pair tail cures the ripple problem but not the distortion. Using high-beta input transistors reduces both problems but does not eliminate them.

# 25 Diagnosing distortions

*January 1998*

This chapter focuses on the visual appearance of the distortion residuals produced by THD analysis. The distortion products may be low-order, in which case they appear simply as second or third harmonic, looking basically sinusoidal.

Crossover distortion, the greatest disturbance in the amplifier designer's peace of mind, is instantly recognizable not only by its shape but by its timing, coinciding as it does with the zero-crossings of the output waveform. The other distortions are slightly harder to distinguish, as they all have their origin in contamination of the signal with half-wave rectified sinewaves, and so the different residuals look similar. Fortunately it usually takes only one or two simple experiments to determine the root cause of the trouble. On several occasions I have watched people trying to impose their will on a recalcitrant design without looking at the distortion residual, an approach which leaves me shaking my head in bafflement. To me the THD analyser is as much a part of the amplifier designer's armoury as the doctor's stethoscope, the pathologist's microscope or even the geologist's hammer.

In recent years, some audio commentators have been rudely dismissive of the simplest and most basic kind of distortion measurement – the total-harmonic distortion, or THD, test.

Because THD measurement has a long history, it is easy to imply that it is outdated and used only by the clueless. This is not so. Many other distortion tests exist, but none of them allows instant diagnosis of audio problems with one glance at an oscilloscope.

The test requires an oscillator with negligible distortion, feeding the unit under test. A notch-filter then completely suppresses the fundamental, to reveal the distortion products that have been generated. What remains after the fundamental is removed is not unnaturally called the THD residual.

## The blameless amplifier concept

A Blameless amplifier results when the known distortions in the panel on distortion mechanisms have been either minimised or reduced to below visibility on the THD residual. It is so-called because it achieves its superb linearity not by startling innovation but simply by avoiding a series of possible errors. Avoiding them is straightforward once they are identified.

The concept of a Blameless amplifier has proved extremely useful. Such an amplifier has surprisingly low THD, despite its conventional-looking circuitry, but its greatest advantage is its defined performance, only weakly dependent on component characteristics.

If an amplifier does not perform to Blameless standards of linearity, then there is something fairly simple wrong with it, and to attempt to improve it by adding extra circuitry or turning up the bias into Class AB misses the point totally.

In several previous articles I have described the various distortions that afflict audio power amplifiers. In the generic circuit, these are relatively few in number as is evident from the panel entitled 'Distortion mechanisms'. Here I will show what some of these distortion residuals actually look like. Distortions 1, 2, 4 and 8 are not very informative visually, being essentially second or third harmonic, so I have omitted them to make room for the more complex waveforms specific to Class B.

## Making distortion measurements

Total harmonic distortion is the r.m.s. sum of all the distortion components generated by the path under test. It is usually quoted as a percentage of the total signal level.

The r.m.s. calculation – taking the square-root of the sum of the squares of the harmonics – emphasises spiky distortions, but whether this helps to mimic human perception of distortion is unclear. The peak capability of true-r.m.s. circuitry is limited, and this may well lead to under-reading of crossover spikes and such.

I hold that the best method is to observe the residual simultaneously, and time-aligned, with the output sinewave as in Figure 1. If you are testing similar pieces of equipment, then the gain of the oscilloscope's second channel for the residual can be kept at the same setting. This allows linearity to be assessed at a glance.

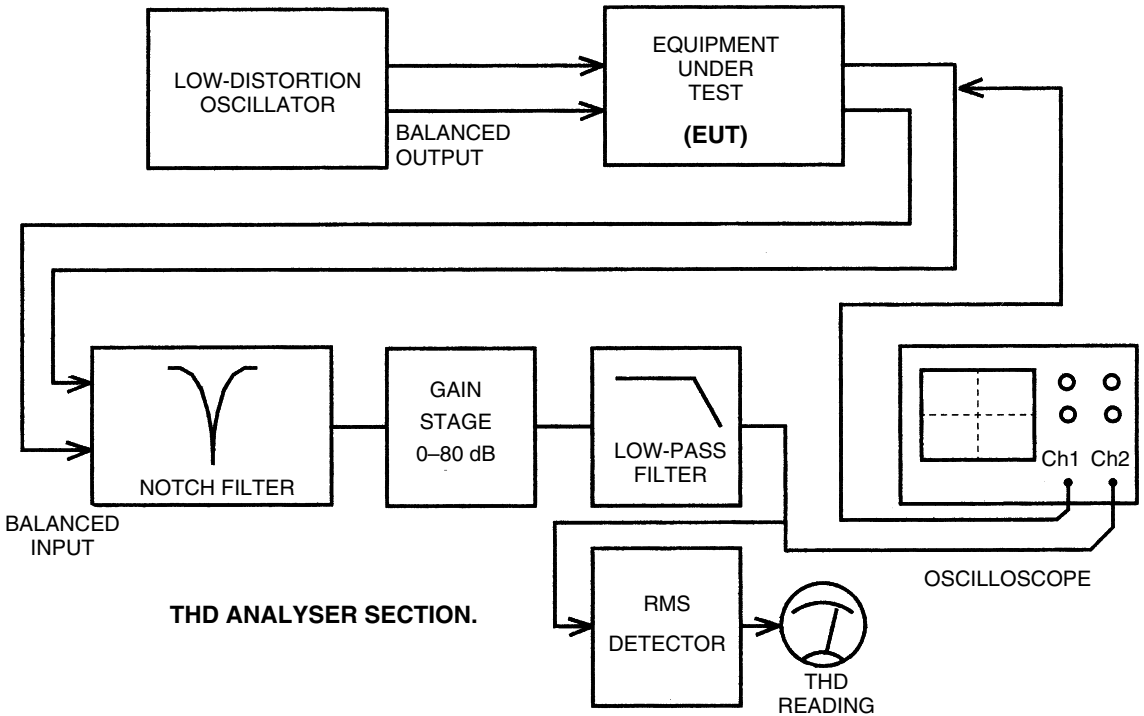



**Figure 1** *Block diagram of a THD analyser. The minimum reading is set by input amplifier noise and oscillator distortion rather than the filter auto-tuning.*

In contrast, it is wiser to connect the actual output to channel 1, rather than an auto-scaled version from the analyser, as this prevents parasitics, etc., from being filtered out by the analyser input circuitry.

The beauty of THD testing is that the error is isolated; in essence, the residual is the difference between perfection and reality. When viewed time-aligned with the output sinewave, crossover distortion can be diagnosed immediately as it occurs at the zero-crossings. On the other hand, non-linearity confined to one peak is probably due to something running out of voltage swing or current capability.

## Two technical challenges

Figure 1 shows the basic THD measuring system. There are two major technical challenges to be overcome. The signal source must be extremely pure, as any oscillator distortion puts an immediate limit on the measurement floor; it must maintain superb performance at least over the range 10 Hz to 20 kHz. A balanced output is highly desirable.

In the analyser section a balanced input is essential. Very great attenuation of the fundamental is required – about 120 dB if you are going to measure down to 0.0005%, making notch tuning is extremely critical. This cannot be attained by fixed-tuned filters, and manual tuning, requiring at least six controls, is about as much fun as picking oakum. In modern THD equipment both frequency and phase are continuously adjusted by a twin servo-loop that optimises the cancellation.

An additional low-pass filter defines the measurement bandwidth. Usually, 80 kHz is a good compromise, retaining most of the important harmonics while reducing noise. A switchable 400 Hz high-pass filter is often fitted, allowing measurements at 1 kHz and up, in the presence of hum. Such a filter should be used only in exceptional cases, for THD often rises sharply at low frequencies, and this would be missed.

While frequently advocated as a more searching examination of an audio path, twin-tone intermodulation tests are almost useless for circuit investigation. They give very little information about the source of the non-linearity as the phase relationship between the test signal and the result is lost. It is often claimed they give a better measure of audible degradation in real use, but a test using two or three tones is still a long way away from music that has tens or hundreds of simultaneous frequencies. Intermodulation tests can often dispense with very-low-THD oscillators, but this in itself is not much of a recommendation.

If real subjective degradation is the issue, a test signal much closer to reality is required. This can be either pseudorandom noise as in the Belcher test,[1] or real music, as in the Baxandall[2] and Hafler[3] cancellation tests.

Returning to harmonic distortion, much better correlation between THD measurements and subjective impairment is possible if the harmonics are weighted so that the higher order components are emphasised.

Weighting by $n^2/4$, so that the second harmonic is unchanged, the third increased by 9/4, and so on, is generally accepted to be roughly correct.[4,5] I was surprised to find that this approach goes back to 1937 and before.[6] I doubt however whether this can be applied to crossover distortion.

When the THD residual is displayed on an analogue oscilloscope, artifacts in the noise are easily detectable by the averaging processes of our vision, but they remain unavailable to conventional measurement. A digital scope can perform even more effective averaging by computation, making submerged distortion artifacts both visually clearer and readily measurable, though an r.m.s. mode may not be available.

If a noisy signal is averaged two times, by combining two sweeps, the coherent signal stays at the same level, while the uncorrelated noise decreases by 3 dB. Averaging 64 times performs this process six-fold, so noise is then reduced by 18 dB. The oscilloscope used here was a digital *HP54600B* 100 MHz digital storage; an excellent instrument. This choice will not come as a surprise to alert readers.

## Distortion mechanisms

My original series on amplifiers[7] listed seven independent distortions inherent to the generic/Lin Class-B amplifier, and whose existence is not dependant on circuit details. I have now increased this to eight.

### *Distortion one*

**Input-stage distortion.** Non-linearity in the input stage. If this is a carefully-balanced differential pair then the distortion is typically only measurable at high frequencies, rises at 18 dB/octave, and is almost pure third harmonic.

If the input pair is unbalanced – which from published circuitry it usually is – then enough second harmonic is produced to swamp the third. Hence the h.f. distortion emerges from the noise at a lower frequency, rising at 12 dB/octave.

### *Distortion two*

**Voltage amplifier stage distortion.** Surprisingly, non-linearity in the voltage-amplifier stage does not always contribute significantly in the total distortion. If it does, it remains constant until the dominant-pole frequency P1 is reached, and then rises at 6 dB/octave. In the generic configuration discussed here it is always second harmonic.

### *Distortion three*

**Output-stage distortion.** Non-linearity in the output stage – the most obvious source. This has three components: crossover distortion (3a) usually dominates for Class-B into 8 Ω, generating high-order harmonics rising at 6 dB/octave as global negative feedback decreases. Low-order large-signal nonlinearity (3b) appears with 4 Ω loads and worsens at 2 Ω. Distortion 3c stems from overlap of output device conduction and only appears at high frequencies.

### *Distortion four*

**Voltage-amplifier loading.** Loading of the voltage-amplifier stage by the non-linear input impedance of the output stage.

### *Distortion five*

**Rail decoupling distortion.** Non-linearity caused by large rail-decoupling capacitors feeding the distorted signals on the supply lines into the signal ground. This seems to be the reason that many amplifiers have rising THD at low frequencies.

### *Distortion six*

**Induction distortion.** Induction of Class-B supply currents into the output, ground, or negative-feedback lines. Almost certainly the least understood and so must common distortion afflicting commercial amplifiers.

### *Distortion seven*

**Negative-feedback take-off distortion.** Non-linearity resulting from taking the negative feedback feed from slightly the wrong place near the point where Class-B currents sum to form the output.

### *Distortion eight*

Capacitor distortion. Rising as frequency falls, capacitor distortion is caused by non-linearity in the input d.c.-blocking capacitor or the feedback network capacitor. The latter is more likely.

### *Distortions x*

**Non existent or negligible distortions.** Common-mode distortion in the input stage and thermal distortion in the output stage – or anywhere else.

Although sometimes invaluable, digital oscilloscopes are often not the best choice for audio THD testing and general amplifier work; in particular the problems of aliasing make the detection and cure of h.f. oscillations very difficult.

To create the residuals shown here, a Blameless amplifier was used essentially identical to that published in Ref. 7. Output was 25 W into 8 Ω, or 50 W into 4 Ω. The Blameless amplifier concept is outlined in a separate panel.

## Crossover distortion

Crossover distortion is only one of the three components that make up Distortion 3 but is often the dominant one. Blameless amplifiers show only crossover distortion when driving 8 Ω or more, and at low and medium frequencies it should be below the noise. This remains true even if the amplifier noise is within a few decibels of the theoretical minimum from a 50 Ω source resistance.

Figure 2 shows the THD residual from such a Blameless power amplifier, with optimally biased in Class-B. Since this is a record of a single sweep, the residual appears to be almost wholly noise. The visual averaging process

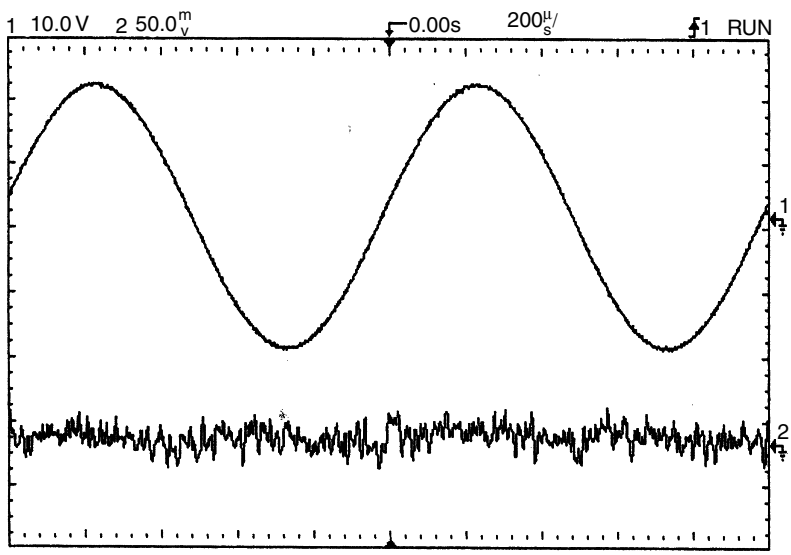



**Figure 2** *The THD residual from an optimally-biased Blameless power amplifier at* 1 kHz, 25 W/8 Ω *is essentially white noise. There is some evidence of artifacts at the crossover point, but they are not measurable. THD* 0.00097%, 80 kHz *bandwidth.*

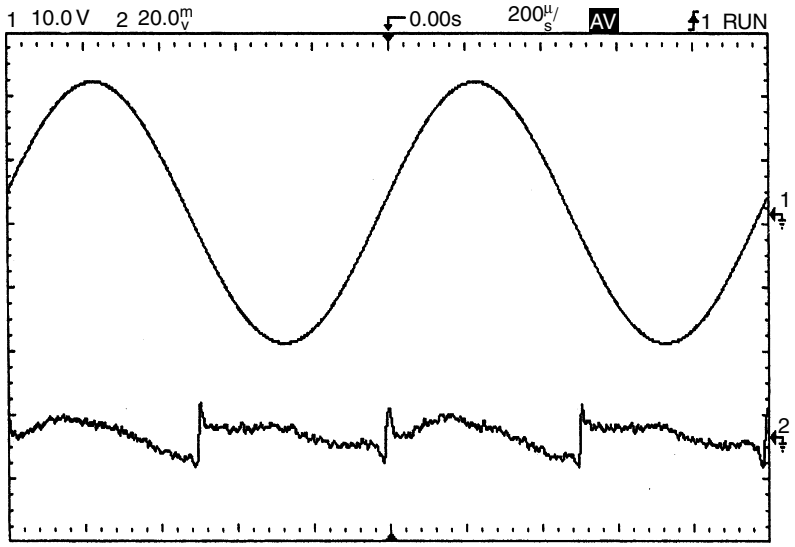



**Figure 3** *Averaging the Figure 2 residual 64 times reduces the noise by* 18 dB, *and crossover discontinuities are now obvious. The residual has been scaled up by 2.5 times from Figure 2 for greater clarity.*

is absent and so the crossover artifacts are actually less visible than on an analogue scope in real time.

In Figure 3, 64 times digital averaging is applied, which makes the disturbances around crossover very clear. A low-order component at roughly 0.0003% is also revealed, which is probably due to very small amounts of Distortion 6 that were not visible when the amplifier layout was optimised.

Figure 4 shows Class B mildly underbiased to generate crossover distortion. The crossover spikes are very sharp, and their height in the residual depends critically on measurement bandwidth. Their presence warns immediately of underbiasing and avoidable crossover distortion.

In Figure 5 an optimally-biased amplifier is tested at 10 kHz. The THD has increased to approx 0.004%, as the amount of global negative-feedback is 20 dB less than at 1 kHz. The crossover events appear wider than in Figure 3. The higher THD level is above the noise so the residual is averaged eight times only.

The measurement bandwidth is still 80 kHz, so harmonics above the eighth are lost. This is illustrated in Figure 6, which is Figure 5 rerun with a 500 kHz bandwidth. The distortion products look very different.

The 80 kHz cutoff point is something of *de facto* standard, which is reasonable as it seems highly unlikely that ultrasonic harmonics can detract from one's listening pleasure. This does not mean THD testing can stop at 10 kHz, as there might be an area of bad intermodulation in the top octave.

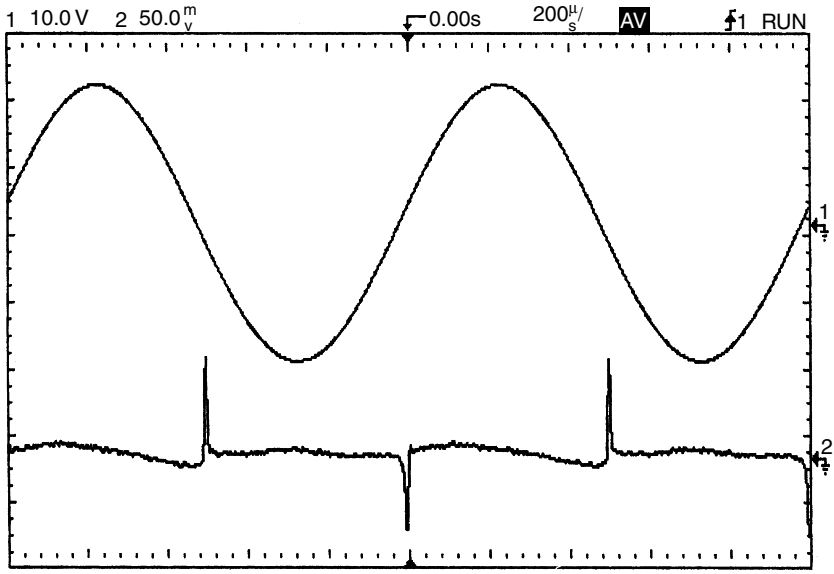



**Figure 4** *Results of mild underbias in Class B.*

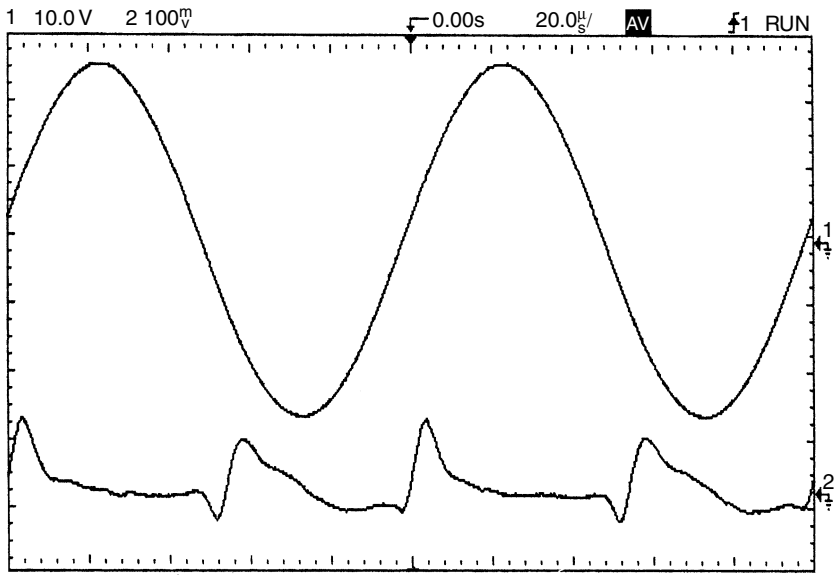



**Figure 5** *An optimally-biased Blameless power amplifier at* 10 kHz. *THD is around 0.004%, bandwidth* 80 kHz. *Averaged eight times.*

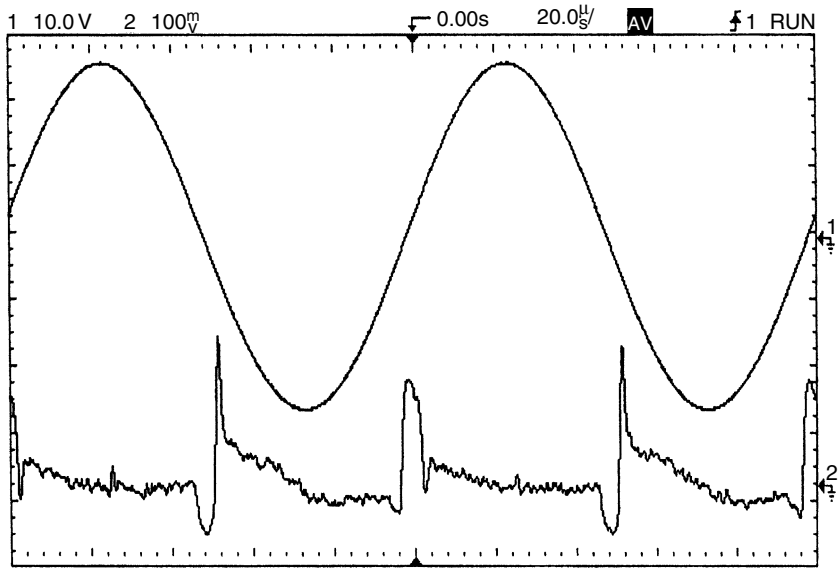



**Figure 6** *As Figure 6, but in* 500 kHz *bandwidth. The distortion products look quite different.*

My practice is to test up to 50 kHz, to check that nothing awful is lurking just outside the audio band; this is safe for moderate powers, and short durations.

## Classes B and AB

I showed in my series on power amplifier distortion[7] that Class AB is not a true compromise between Class A and Class B operation. If AB is used to trade off efficiency and linearity, its linearity is superior to B since below the AB transition level, it is pure Class A.

The Class-A region can – and should – have very low THD indeed, below 0.0006% up to 10 kHz, as demonstrated in Ref. 8. However, above the AB transition level THD abruptly worsens. This is due to what has been called '$g_m$-doubling', but is better regarded as a step in the gain/output-voltage relationship. Linearity is then inferior not only to Class-A but also to optimal-bias Class-B.

It is possible to make Class AB distortion very low by proper design. Basically, this means using the lowest possible emitter resistors to reduce the size of the gain step.[9] Even so, THD remains at least twice as high as Class-B.

Tweaking up the bias of a Class-B amplifier most certainly does *not* offer a simple trade-off between power dissipation and overall linearity, despite the constant repetition this notion receives in some parts of the audio

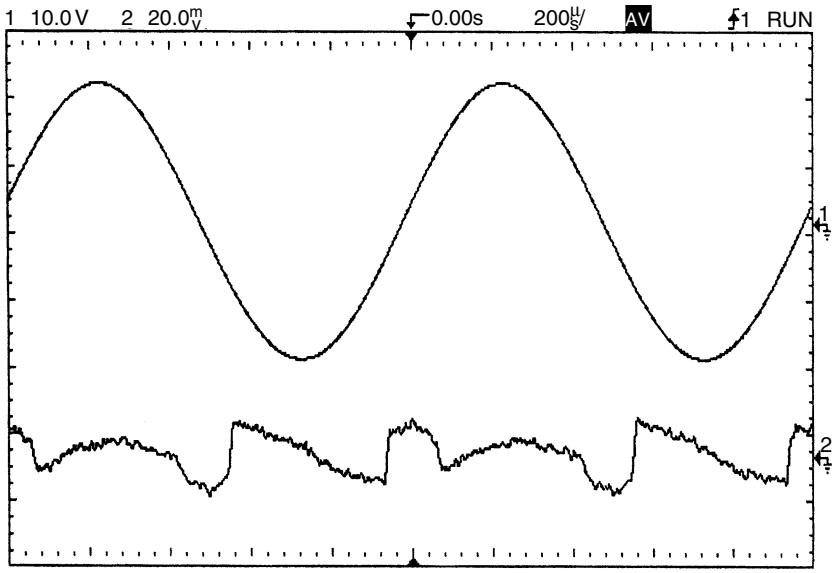



**Figure 7** *The $g_m$-doubling distortion introduced by Class AB. The edges in the residual are larger and no longer at the zero-crossing, but displaced either side of it.*

press. The real choice is: very low THD at low power and high THD at high power, or medium THD at all powers. The electricity bill is another issue.

Figure 7 shows the gain-step distortion introduced by Class AB. The undesirable edges are caused by gain changes that are no longer partially cancelled at the crossover; they are now displaced to either side of the zero-crossing. No averaging is used here as the THD is higher and well above the noise.

## Large-signal non-linearity

When the load resistance falls below 8 Ω, extra low-order distortion components appear. This is true for most or all modern power bipolar junction transistors, but with old devices like 2N3055, some large-signal nonlinearity may appear at 8 Ω. This is a compressive non-linearity, i.e. gain falls as level increases, 'squashing' the signal, and is due to fall-off of transistor beta at high collector currents.

Figure 8 shows the typical appearance of large-signal non-linearity, driving 50 W into 4 Ω, and averaged 64 times. The extra distortion appears to be a mixture of third harmonic, due to the basic symmetry of the output stage, with some second harmonic, because the beta-loss is component-dependant and not perfectly symmetrical in the two halves of the output.

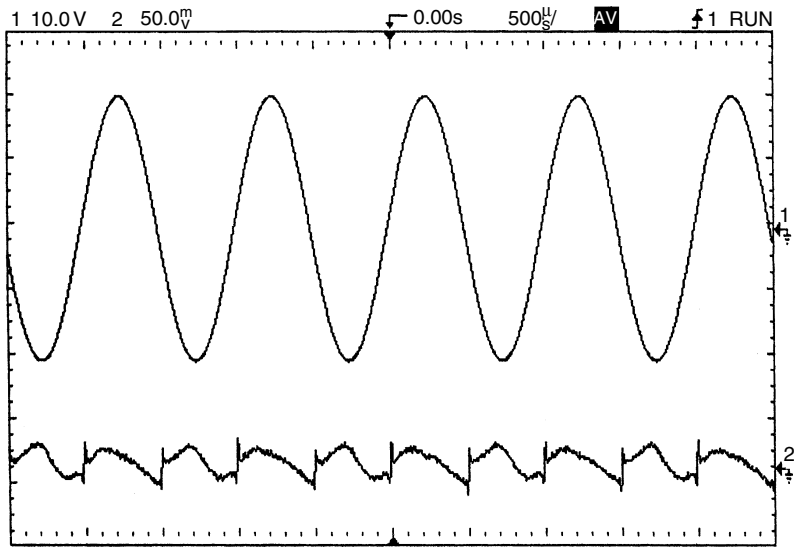



**Figure 8** *Large-signal nonlinearity, driving* 50 W *into* 4 Ω*, and averaged 64 times. The extra distortion appears to be a mixture of third harmonic – occurring as a consequence of the compressive nature of beta-loss – and second harmonic arising because the beta-loss is not perfectly symmetrical in the two halves of the output stage.*

## Other distortions

Of the distortions that afflict generic Class-B power amplifiers, 5, 6 and 7 all look rather similar in the THD residual. This is perhaps not surprising since all result from adding half-wave disturbances to the signal.

Distortion 5 is usually easy to identify as it is accompanied by 100 Hz power-supply ripple; 6 and 7 introduce no ripple. Distortion 6 is easily identified if the d.c. power cables are movable, for altering their run will strongly affect the quantity generated.

Figure 9 shows Distortion 5, provoked by connecting the negative supply rail decoupling capacitor to the input ground instead of giving it its own return to the far side of the star point. Doing this increases THD from 0.00097% to 0.008%, mostly as second harmonic. Ripple contamination is significant and contributes to the THD figure. It could be easily filtered out to make the measurement, but this is just brushing the problem under the carpet.

Distortion 6 is displayed in Figure 10. The negative supply rail was run parallel to the negative-feedback line to produce this diagram. Although more than doubled, THD is still relatively low at 0.0021%, so 64-times averaging is used.

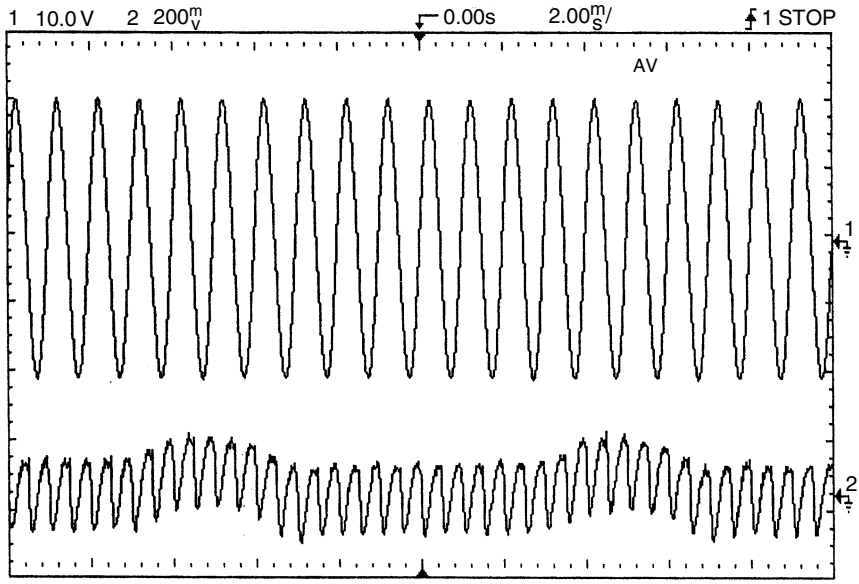



**Figure 9** *Distortion 5 revealed. Connecting the rail decoupler to input ground increases THD eight-fold from* 0.00097% *to* 0.008%, *mostly as second harmonic.* 100 Hz *ripple is also visible. No averaging.*

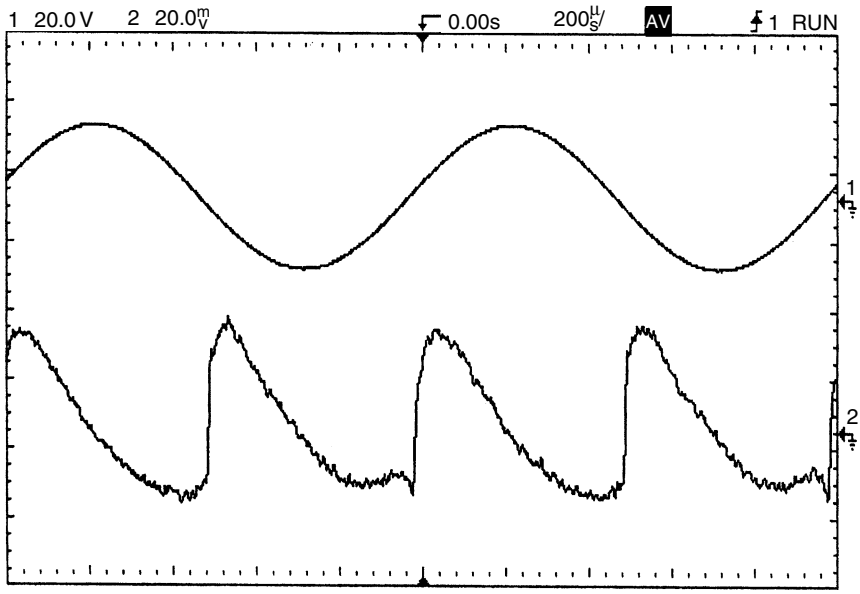



**Figure 10** *Distortion 6. Induction of half-wave signal from the negative supply rail into the negative feedback line increases THD to* 0.0021% *Averaged 64 times.*

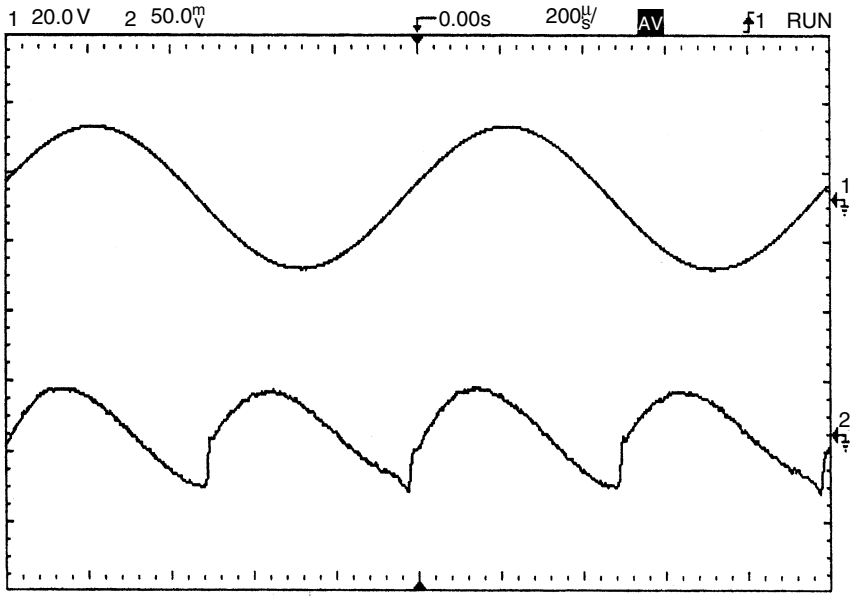



**Figure 11** *Distortion 7, caused by choosing an negative feedback take-off point inside the Class-B output stage rather than on the output line itself. THD increases from* 0.00097% *to* 0.0027%, *by taking the negative feedback from the wrong end of* 10 mm *of very thick resistor leg. Averaged 64 times.*

Figure 11 shows a case of Distortion 7, introduced by deliberately making a minor error in the negative feedback take-off point.

If it is attached to a part of the Class-B output stage so that half-wave currents flow through it, rather than being on the output line itself, THD is increased. Here it rose from 0.00097% to 0.0027%, caused by taking the negative feedback from the wrong end of the leg of one of the output emitter resistors, $R_e$.

Note this was at the right end of the resistor, otherwise THD would have been gross, but 10 mm along a very thick resistor leg from the output line junction. Truly, God is in the details.

## Diagnosis

The rogue's gallery of real-life THD residuals portrayed here will hopefully help with the problem of identifying the distortion mechanism in a misbehaving amplifier. There is no reason why the generic/Lin configuration should give measurable THD at 1 kHz, or more than say, 0.004% at 10 kHz when driving 8 Ω.

It is important to be sure that you are measuring a real distortion mechanism, and not the results of parasitic oscillation upsetting circuit conditions; the oscillation itself may be outside the scope bandwidth. Parasitics usually vary greatly when a cautious finger is applied to the relevant section of the circuitry. Real distortion changes little, though the THD reading will probably be increased by the introduction of hum.

I hope I have shown that THD testing gives an immediate view into circuit operation that other methods do not, however useful they may be in other applications.

It cannot be stated too strongly that to attempt amplifier design and diagnosis without continuous visual observation of the THD residual is to work blind. You will proverbially fall into the ditch.

## References

1. Belcher, R. 'A new distortion measurement', *Wireless World*, May 1978, pp 36–41.
2. Baxandall, P. 'Audible amplifier distortion is not a mystery', *Wireless World*, November 1977, pp 63–66.
3. Hafler, D. 'A listening test for amplifier distortion', *HiFi News & RR*, November 1986, pp 25–29.
4. Moir, J. 'Just detectable distortion levels', *Wireless World*, February 1981, p 34.
5. Shorter, D. 'Influence of high-order products in non-linear, distortion', *Electronic Engineering*, April 1950, p 152.
6. Callendar, M, Letter, *Electronic Engineering*, October 1950, p 443.
7. Self, D. 'Distortion in power amplifiers', *Electronics World*, August 1993 to March 1994.
8. Self, D. 'Distortion in power amplifiers, part 8', *Electronics & Wireless World*, March 1994, p 225 (Class A amp).
9. Self, D. 'Trimodal Audio Power', *Electronics World*, June 1995, p 462. (Effect of $R_e$ in Class AB).

# 26 Trimodal audio power, Part I

*June 1995*

This project had its roots in a large number of requests for a PCB for the Class-A amplifier described in the last part of the *Distortion In Power Amplifiers* series. Rather than just using the original circuit, I took the chance to look closely at output stage voltage efficiency, and also to re-examine the input stage balance and noise performance. I added a safety circuit to prevent catastrophe if the quiescent-current control arrangements went haywire, and this very conveniently doubles as the bias voltage generator when the amplifier is in the Class-B mode.

This proved a very popular design, and I had a lot of correspondence about it. People are still building them.

I present here my own contribution to global warming in the form of an improved Class-A amplifier that I believe is unique. It not only copes with load impedance dips by means of an unusually linear form of Class-AB, but will also operate as a 'blameless' Class-B engine. The power output in pure Class-A is 20 to 30 W into 8 Ω, depending on the exact supply rails chosen.

Initially, I simply intended to provide an updated version of the Class-A circuit published in reference 1, in response to requests for a PCB for the Class-A amplifier designed with my methodology. I decided to use a complementary-feedback-pair (CFP), output stage for best possible linearity, and some incremental improvements have been made to noise, slew rate and maximum d.c. offset.

Naturally, the Class-A circuit bears a very close resemblance to a 'blameless' Class-B amplifier. As a result, I decided to retain the Class-B $V_{be}$ multiplier, and use it as a safety-circuit to prevent catastrophe if the relatively complex Class-A current-regulator failed. From this the idea arose of making the amplifier instantly switchable between Class-A/AB and Class-B modes. This gives two kinds of amplifier for the price of one, and permits of some interesting listening tests. Now you really can do an A/B comparison.

In the Class-B mode the amplifier has the usual negligible quiescent dissipation, but in Class-A the thermal efflux is naturally considerable. This is because true Class-A operation is extended down to 6 Ω resistive loads for the full output voltage swing, by suitable choice of the quiescent current.

With heavier loading the amplifier gracefully enters Class-AB, in which it will give full output down to 3 Ω before the safe-operating-area (SOAR), limiting begins to act. Output into 2 Ω is severely curtailed, as it must be with only one output pair, and this kind of load is not advisable.

In short, the amplifier allows a choice between being firstly very linear all the time – blameless Class-B – and secondly ultra-linear most of the time – Class-A – with occasional excursions into Class-AB.

The amplifier's AB mode is still extremely linear by current standards, though inherently it can never be as good as properly-handled Class-B, and nothing like as good as A. Since there are three possible classes of operation I have decided to call the design a Trimodal power amplifier. It is impossible to be sure that you have read all the literature on an area of technology; however, to the best of my knowledge this is the first ever Trimodal amplifier.

As I said earlier, designing a low-distortion Class-A amplifier is in general a good deal simpler than the same exercise for Class-B. All the difficulties of arranging the best possible crossover between the output devices disappear. Because of this it is hard to define exactly what 'blameless' means for a Class-A amplifier.

In Class-B the situation is quite different, and 'blameless' has a very specific meaning; when each of the eight or more distortion mechanisms has been minimised in effect, there always remains the crossover distortion inherent in Class-B. There appears to be no way to reduce it without departing radically from that might be called the generic Lin amplifier configuration. Therefore the 'blameless' state appears to represent some sort of theoretical limit for Class-B, but not for Class-A.

However, Class-B considerations cannot be ignored, even in a design intended to be Class-A only, because if the amplifier does find itself driving a lower load impedance than expected, it will move into Class-AB. In this case, all the additional Class-B requirements are just as significant as for a Class-B design proper. Class-AB can never give distortion as low as optimally-biased Class-B, but it can be made comparable if the extra distortion mechanisms are correctly handled.

My correspondence has made it abundantly clear that *EW* readers are not going to be satisfied with anything less than state-of-the-art linearity, and so the amplifier described here uses the CFP type of output stage, which has the lowest distortion due to the local feedback loops enclosing the output devices. It also has the advantage of better output efficiency than the emitter-follower version, and inherently superior quiescent current

stability. It will shortly be seen that these are both important for this design.

Half-serious thought was given to labelling the Class-A mode 'distortionless' as the THD is completely unmeasurable across most of the audio band. However, detectable distortion products do exist above 10 kHz, so sadly, I abandoned this provocative idea.

Before putting cursor to CAD, it seemed appropriate to take another look at the Class-A design, to see if it could be inched a few steps nearer perfection. The result is a slight improvement in efficiency, and a 2 dB improvement in noise performance. In addition the expected range of output d.c. offset has been reduced from ±50 mV to ±15 mV, still without any adjustment.

## The power and the glory

The amplifier is 4 Ω capable in both A/AB and B operating modes, though it is the nature of things that the distortion performance is not quite so good. All solid-state amplifiers – without qualification, as far as I am aware – are much happier with an 8 Ω load, both in terms of linearity and efficiency; loudspeaker designers please note.

With a 4 Ω load, Class-B operation gives better THD than Class-A/AB, because the latter will always be in AB mode, and therefore generating extra output stage distortion through $g_m$-doubling. This should really be called gain-deficit-halving, but somehow I don't see this term catching on. These not entirely obvious relationships are summarised on the right.

Figure 1 attempts to show diagrammatically just how power, load resistance, and operating mode are related. The rails have been set to ±20 V, which just allows 20 W into 8 Ω in Class-A. The curves are lines of constant power, i.e. $V \times I$ in the load, the upper horizontal line represents maximum voltage output, allowing for $V_{ce(sat)}{}^S$, and the sloping line on the right is the SOAR protection locus; the output can never move outside this area in either mode. The intersection between the load resistance lines sloping up from the origin and the ultimate limits of voltage-clip and SOAR protection define which of the curved constant-power lines is reached.

In A/AB mode, the operating point must be left of the vertical push-pull current-limit line (at 3A, i.e. twice the quiescent current) for Class-A. If we move along one of the impedance lines, when we pass to the right of the push-pull limit the output devices will begin turning off for part of the cycle; this is the AB operation zone. In Class-B mode, the 3A line has no significance and the amplifier remains in optimal Class-B until clipping or SOAR limiting occurs. Note that the diagram axes represent instantaneous power in the load, but the curves show sine-wave r.m.s. power, and that is the reason for the apparent factor-of-two discrepancy between them.

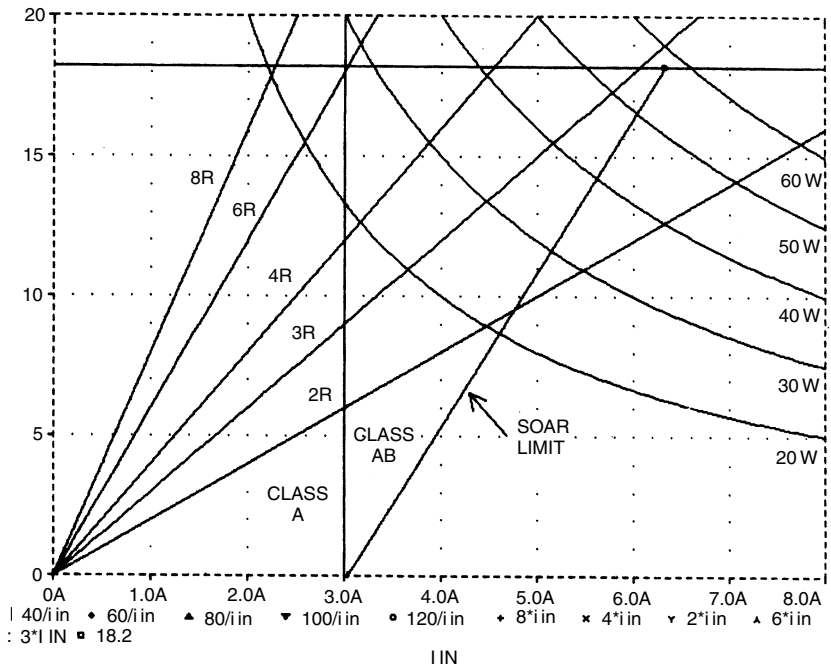



**Figure 1** *Relationships between load, mode, and power output. The intersection between the sloping load resistance lines and the ultimate limits of voltage-clipping and SOAR protection define which of the curved constant-power lines is reached. In A/AB mode, the operating point must be to the left of the vertical push-pull current-limit line for true Class-A.*

**Table 1**

| *Load* (Ω) | *Mode* | *Distortion* | *Dissipation* |
|---|---|---|---|
| 8 | A/AB | very low | high |
| 4 | A/AB | high | high |
| 8 | B | low | low |
| 4 | B | medium | medium |

Note that in the context of this sort of amplifier, 'high' means about 0.002% THD at 1 kHz and 0.01% at 10 kHz.

## Health and efficiency

Concern for efficiency in Class-A may seem paradoxical, but one way of looking at it is that Class-A watts are precious things, wrought in great heat and dissipation, and so for a given quiescent power it makes sense to ensure that the amplifier approaches its limited theoretical efficiency as

closely as possible. I was confirmed in this course by reading of another recent design[2] which seems to throw efficiency to the winds by using a hybrid bjt/FET cascode output stage. The voltage losses inherent in this arrangement demand ±50 V rails and sixfold output devices for a 100 W Class-A capability; such rail voltages would give 156 W from a 100% efficient amplifier.

Voltage efficiency of a power amplifier is the fraction of the supply-rail voltage which can actually be delivered as peak-to-peak voltage swing into a specified load; efficiency is invariably less into 4 Ω due to the greater resistive voltage drops with increased current.

The Class-B amplifier I described in Ref. 3 has a voltage efficiency of 91.7% for positive swings, and 92.5% for negative, into 8 Ω. Amplifiers are not in general completely symmetrical, and so two figures need to be quoted; alternatively the lower of the two can be given as this defines the maximum undistorted sine-wave. These figures above are for an emitter-follower output stage, and a complementary-feedback pair output does better, the positive and negative efficiencies being 94.0% and 94.7% respectively.

The emitter follower version gives a lower output swing because it has two more $V_{be}$ drops in series to be accommodated between the supply rails; the CFP is always more voltage-efficient, and so selecting it over the emitter follower for the current Class-A design is the first step in maximising efficiency.

Figure 2 shows the basic complementary-feedback pair output stage, together with its two biasing elements. In Class-A the quiescent current is rigidly controlled by negative-feedback; this is possible because in Class-A the total voltage across both emitter resistors $R_e$ is constant throughout the cycle. In Class-B this is not the case, and we must rely on 'thermal feedback' from the output stage, though to be strictly accurate this is not 'feedback' at all, but a kind of feed-forward.

It is a big advantage of the CFP configuration that quiescent current, $I_q$ depends only on driver temperature, and this is important in the Class-B mode, where true feedback control of quiescent current is not possible. This has special force if low-value emitter resistors such as 0.1 Ω, are chosen, rather than the more usual 0.22 Ω; the motivation for doing this will soon become clear.

Voltage efficiency for the quasi-complementary Class-A circuit of Ref. 1 into 8 Ω is 89.8% positive and 92.2% negative. Converting this to the CFP output stage increases this to 92.9% positive and 93.6% negative. Note that a Class-A $I_q$ of 1.5 A is assumed throughout; this allows 31 W into 8 Ω in push-pull, if the supply rails are adequately high. However the assumption that loudspeaker impedance never drops below 8 Ω is distinctly doubtful, to put it mildly, and so as before this design allows for full Class-A output voltage swing into loads down to 6 Ω.

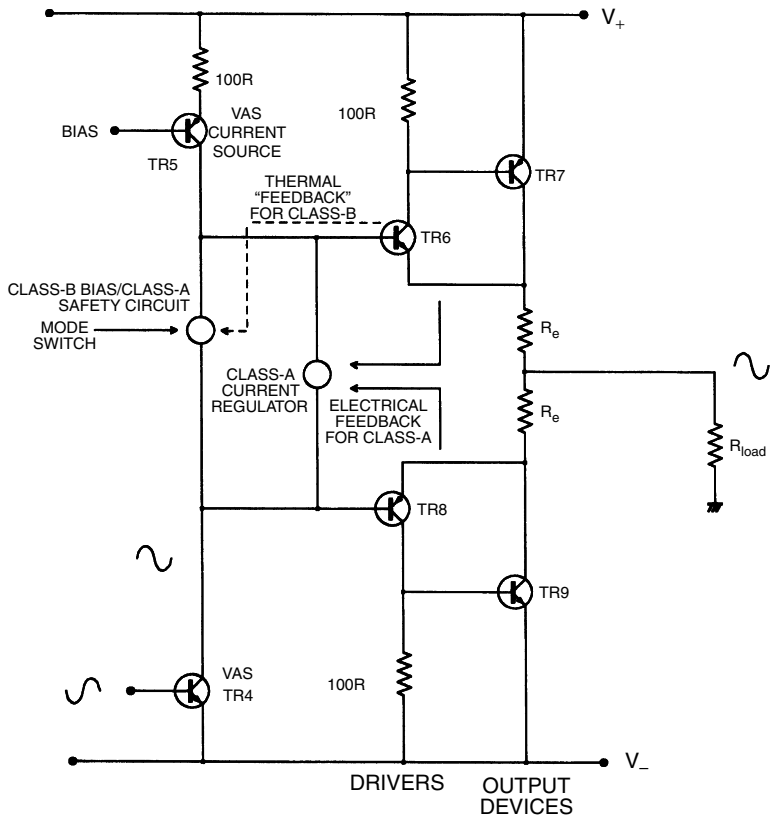



**Figure 2** *Basic current feedback output stage, equally suited to operating Class B, AB and A, depending the magnitude of* $V_{bias}$. *The resistors* $R_e$ *may be from* 0.1 *to* 0.47 Ω.

So how else can we improve efficiency? The addition of extra and higher supply rails for the small-signal section of the amplifier surprisingly does not give a significant increase in output; examination of Figure 3 shows why. In this region of operation, the output device $Tr^7$ base is at a virtually constant 880 mV below the positive rail, and as $Tr_6$ driver base rises it passes this level, and keeps going up; clipping has not yet occurred.

The driver emitter follows the driver base up, until the voltage difference between this emitter and the output base, i.e. the driver $V_{ce}$, becomes too small to allow further conduction; this choke point is indicated by the arrows A-A. At this point the driver base is forced to level off, although it is still about 500 mV below the level of the positive rail. Note also how the voltage between the positive rail and $Tr_5$ emitter collapses. Thus a higher rail will give no extra voltage swing, which I must admit came as something of a surprise. Higher sub-rails for small-signal sections only come into their

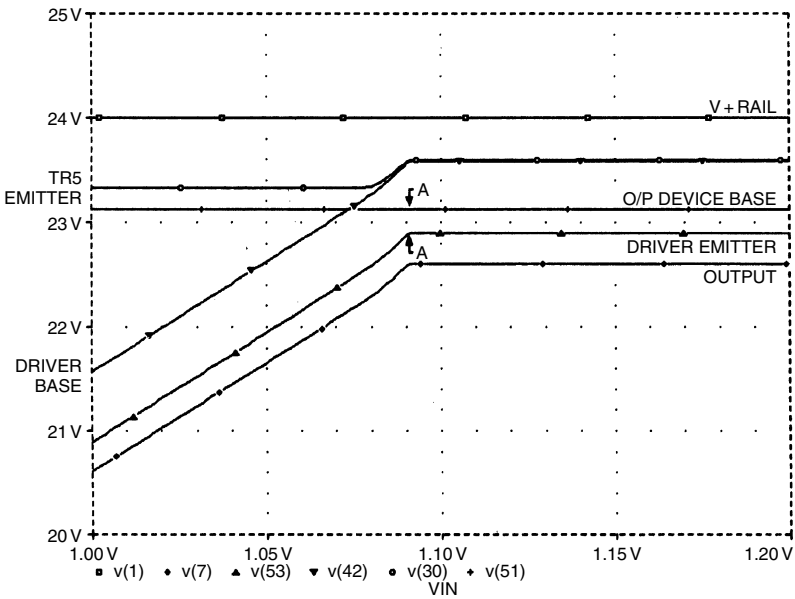



**Figure 3** *PSpice simulation showing how positive clipping occurs in the current feedback output. A higher sub-rail for the voltage amplifier cannot increase the output swing, as the limit is set by the minimum driver $V_{ce}$ and not the voltage amplifier output swing.*

own in FET amplifiers, where the high $V_{gs}$ for FET conduction (5 V or more) makes their use almost mandatory.

Efficiency figures given so far are all greater for negative rather than positive voltage swings. The approach to the rail for negative clipping is slightly closer because there is no equivalent to the 0.6 V bias established across $R_{13}$; however this advantage is absorbed by the need to lose a little voltage in the *RC* filtering of the negative supply to the current-mirror and voltage amplifier stage. This filtering is essential if really good ripple/hum performance is to be obtained[3].

In the quest for efficiency, an obvious variable is the value of the output emitter resistors $R_e$. The performance of the current-regulator described, especially when combined with a CFP output stage, is more than good enough to allow these resistors to be reduced while retaining first-class $I_q$ stability. I took 0.1 Ω as the lowest practicable value, and even this is comparable with PCB track resistance, so some care in the exact details of physical layout is essential; in particular the emitter resistors must be treated as four-terminal components to exclude unwanted voltage drops in the tracks leading to the resistor pads.

If $R_e$ is reduced from 0.22 Ω to 0.1 Ω then voltage efficiency improves from 92.9%/93.6%, to 94.2%/95.0%. Is this improvement worth having?

Well, the voltage-limited power output into 8 Ω is increased from 31.2 to 32.2 W with ±24 V rails, at absolutely zero cost, but it would be idle to pretend that the resulting increase in sound-pressure level is highly significant. It does however provide the philosophical satisfaction faction that as much Class-A power as possible is being produced for a given dissipation; a delicate pleasure.

The linearity of the CFP output stage in Class-A is very slightly worse with 0.1 Ω emitter resistors, though the difference is small and only detectable open-loop; the simulated THD of an output stage alone (for 20 V pk-pk in 8 Ω) is only increased from 0.0027% to 0.0029%. This is probably due to simply to the slightly lower total resistance seen by the output stage.

However, at the same time reducing the emitter resistors to 0.1 Ω provides much lower distortion when the amplifier runs out of Class-A; it halves the size of the step gain changes inherent in Class-AB, and so effectively reduces distortion into 4 Ω loads.

Figures 4 & 5 are output linearity simulations; the measured results from a real and 'blameless' Trimodal amplifier are shown in Figure 6, where it can be clearly seen that THD has been halved by this simple change. To the best of my knowledge this is a new result; my conclusion is that if you must work in Class-AB, keep the emitter resistors as low as possible, to minimise the gain changes.

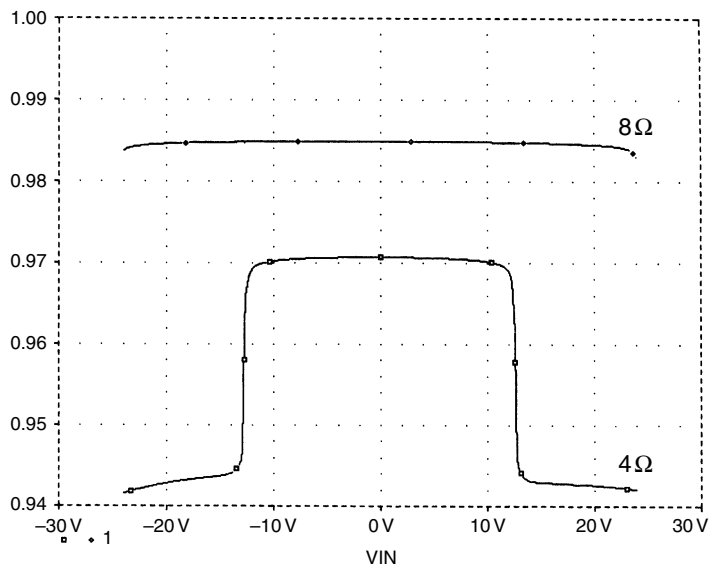



**Figure 4** *Complementary feedback pair output stage linearity with $R_e$ set at 0.22 Ω. Upper trace is Class-A into 8 Ω, lower is Class-AB operation into 4 Ω, showing step changes in gain of 0.024 units.*

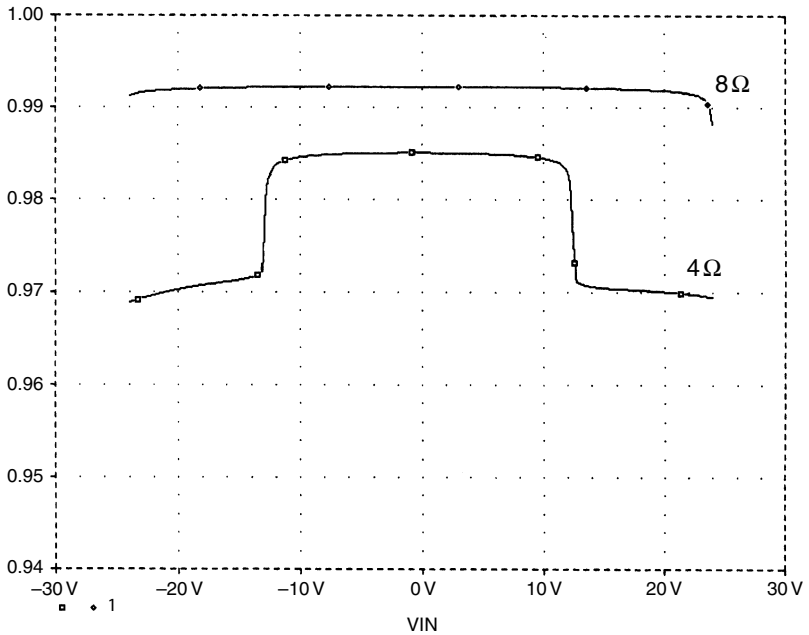



**Figure 5** *Current feedback output linearity with $R_e$ set at* 0.1 Ω*, re-biased to keep $I_q$ at* 1.5 A*. There is slightly poorer linearity in the flat-topped Class-A region than for an $R_e$ of* 0.22 Ω*, but the* 4 Ω *AB steps are halved in size at .012 units. Note that both gains are now closer to unity; same scale as Figure 4.*

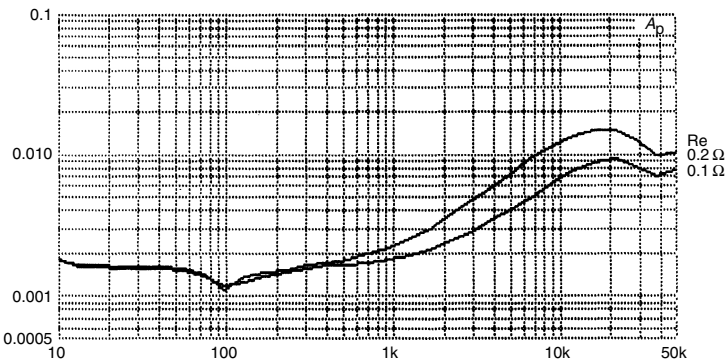



**Figure 6** *Proving that emitter resistor value really matters in Class-AB, Output was* 20 W *in* 4 Ω*, so amplifier was leaving Class-A for about* 50% *of the time. Changing emitter resistors from* 0.2 *to* 0.1 Ω *halves the distortion. Current $I_q$ is* 1.5 A *for both cases.*

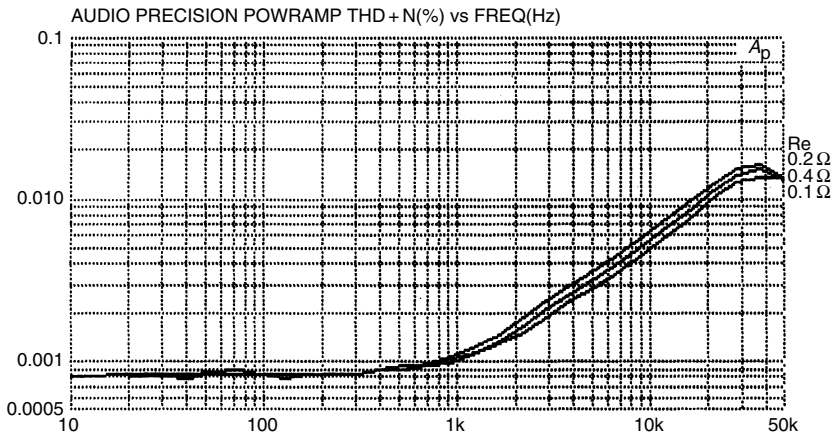



**Figure 7** *Proving that emitter resistors matter much less in Class-B. Output was* 20 W *in* 8 Ω*, with optimal bias. Interestingly, the bias does NOT need adjusting as the value of* $R_e$ *changes. Bandwidth* 80 kHz.

Having considered the linearity of Class-A and AB, we must not neglect what effect this radical Re change has on Class-B linearity. The answer is, not very much, but there is a slightly reduction in THD, Figure 7, where crossover distortion seems to be slightly higher with $R_e$ at 0.2 Ω than for either 0.1 or 0.4 Ω. Whether this is a consistent effect – for complementary-feedback pair stages anyway – remains to be seen.

The detailed mechanisms of bias control and mode-switching are described in the second part of this article.

## Improving noise performance

In a power amplifier, noise performance is not an irrelevance.[4] It is well worth examining just how good it can be. As in most amplifiers, noise is set here by a combination of the active devices at the input and the surrounding resistances.

Operating conditions of the input transistors themselves are set by the demands of linearity and slew-rate, and there is little freedom of design here; however the collector currents are already high enough to give near-optimal noise figures with the low source impedances – a few hundred ohms – that we have here, so this is not too great a problem. Also remember that noise figure is a weak function of $I_c$, so minor tweaking makes no detectable difference. We certainly have the choice of input device type; there are many more possibilities now that we have relatively low rail

voltages. Noise performance is, however, closely bound up with source impedance, and we need to define this before device selection.

Looking therefore to the passives, there are several resistances generating Johnson noise in the input, and the only way to reduce this noise is to reduce them in value. The obvious candidates are input stage degeneration resistors $R_{2,3}$ and $R_9$, which determines the output impedance of the negative-feedback network. There is also another unseen component; the source resistance of the preamplifier or whatever upstream.

Even if this equipment were miraculously noise-free, its output resistance would still generate Johnson noise. If the preamplifier had, say, a 20 kΩ volume pot at its output – not a good idea, as this gives a poor gain structure and cable dependent h.f. losses, but that is another story[5] – then the source resistance could be a maximum of 5 kΩ, which would almost certainly generate enough Johnson Noise to dominate the power-amplifier's noise behaviour. However, there is nothing that power-amp designers can do about this, so we must content ourselves with minimising the noise-generating resistances we do have control over.

The presence of input degeneration resistors $R_{2,3}$ is the price we pay for linearising the input stage by running it at a high current, and then bringing its transconductance down to a useable value by adding linearising local negative feedback. These resistors cannot be reduced, for if the h.f. negative-feed-back factor is then to remain constant, $C_{\mathrm{dom}}$ would have to be proportionally increaseed, with a consequent reduction in slew rate. Used with the original negative feedback network, these resistors degrade the noise performance by 1.7 dB. Like all the other noise measurements given here, this figure assumes a 50 Ω external source resistance.

If we cannot alter the input degeneration resistors, then the only course left is the reduction of the feedback network impedance, and this sets off a whole train of consequences. If $R_8$ is reduced to 2.2 kΩ, then $R_9$ becomes 110 Ω, and this reduces noise output from −93.5 dBu to −95.4 dBu. Note that if $R_{2,3}$ were not present, the respective figures would be −95.2 and −98.2 dBu. However, $R_1$ must also be reduced to 2.2 kΩ to maintain d.c. balance, and this is too low an input impedance for direct connection to the outside world.

If we accept that the basic amplifier will have a low input impedance, there are two ways to deal with it. The simplest is to decide that a balanced line input is essential; this puts an opamp stage before the amplifier proper, buffers the low input impedance, and can provide a fixed source impedance to allow the high and low-frequency bandwidths to be properly defined by an *RC* network using non-electrolytic capacitors. The common practice of slapping an *RC* network on an unbuffered amplifier input must be roundly condemned as the source impedance is unknown, and so therefore is the roll-off point. A major stumbling block for subjectivist reviewing, one would have thought.

The other approach is to have a low resistance d.c. path at the input but maintain a high a.c. impedance; in other words to use the fine old practice of input bootstrapping. Now this requires a low-impedance unity-gain-with-respect-to-input point to drive the bootstrap capacitor, and the only one available is at the amplifier inverting input, i.e. the base of $Tr_3$. While this node has historically been used for the purpose of input bootstrapping[6] it has only been done with simple circuitry employing very low feedback factors.

There is good reason to fear that any monkey business with the feedback point, at $Tr_3$'s base, will add shunt capacitance, creating a feedback pole that will degrade h.f. stability. There is also the awkward question of what will happen if the input is left open-circuit.

Figure 8 shows how the input can be safely bootstrapped.

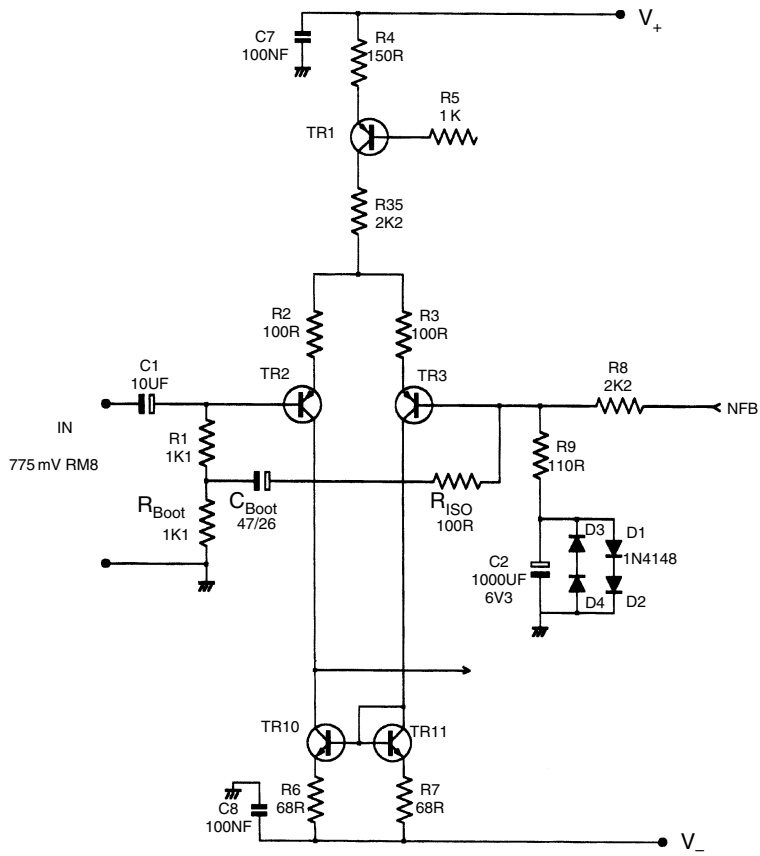



**Figure 8** *Method used for input bootstrapping from the feedback point.* $R_{iso}$ *is essential for dependable high-frequency stability; with it set to* 100 Ω*, input impedance is* 13 kΩ.

The total d.c. resistance of $R_1$ and $R_{boot}$ equals $R_8$, and their centre point is driven by $C_{boot}$. Connecting $C_{boot}$ directly to the feedback point did not produce gross instability, but it did seem to increase susceptibility to sporadic parasitic oscillation. Resistor $R_{iso}$ was added to isolate the feedback point from stray capacitance: this seemed to effect a complete cure.

The input could be left open-circuit without any apparent ill-effects, though this is not exactly good practice if loud-speakers are connected. A value for $R_{iso}$ of 220 Ω increases the input impedance to 7.5 kΩ, and 100 Ω raises it to 13.3 kΩ, safely above the 10 kΩ standard value for a bridging impedance. Despite successful tests, I must admit to a few lingering doubts about the high-frequency stability of this approach, and it might be as well to consider it as experimental until more experience is gained.

Another consequence of a low-impedance negative feedback network is the need for feedback capacitor $C_2$ to be proportionally increased to maintain the low-frequency response, and prevent capacitor distortion from causing a rise in THD at low frequencies; it is the latter constraint that determines the value. This is a separate distortion mechanism from the seven previously considered, and I think deserves the title Distortion 8. This criterion gives a value of 1000 μF, which necessitates a low rated voltage such as 6.3 V if the component is to be of reasonable size. As a result, $C_2$ now needs protective shunt diodes in both directions, because if the amplifier fails it may saturate in either direction.

Close examination of the distortion residual shows that the onset of conduction of back-to-back diodes will cause a minor increase in THD at 10 Hz, from less than 0.001% to 0.002%, even at the low power of 20 W/8 Ω. It is not my practice to tolerate such gross non-linearity, and therefore four diodes are used in the final circuit, and this eliminates the distortion effect, Figure 8. It could be argued that a possible reverse-bias of 1.2 V does not protect $C_2$ very well, but at least there will be no explosion.

We can now consider alternative input devices to the *MPSA56*, which was never intended as a low-noise device. Several high-beta low-noise types such as *2SA970* give an improvement of about 1.8 dB with the low-impedance negative feedback network. Specialised low-$R_b$ devices like *2SB737* give little further advantage – possibly 0.1 dB – and it is probably better to go for one of the high-beta types; the reason why will soon emerge.

It could be argued that the complications of a low-impedance negative feedback network are a high price to pay for a noise reduction of some 2 dB; however, there is a countervailing advantage, for the above negative feedback network modification significantly improves the output d.c. offset performance. The second and final part of this article shows how, and also gives full details of the mode-switching and bias control systems, and the performance of the complete amplifier.

## References

1. Self, D. 'Distortion in power amplifiers', Part 8, *Electronics World & Wireless World*, March 1994, p 225.
2. Thagard, N. 'Build a 100 W Class-A Mono Amp', *Audio*, January 1995, p 43.
3. Self, D. 'Distortion in power amplifiers', Part 7, *Electronics World & Wireless World*, February 1994, p 137.
4. Self, D. 'Distortion in power amplifiers', Part 2, *Electronics World & Wireless World*, September 1993, p 736.
5. Self, D. 'A Precision Preamplifier', *Wireless World*, October 1983, p 31.
6. Mullard Ltd, *Transistor Audio & Radio Circuits*, Mullard Ltd. 1972, second edn., p 122.

# 27 Trimodal audio power, Part II

*July 1995*

The same components that dominate amplifier noise performance also determine the output d.c. offset; if $R_9$ is reduced to minimise the source resistance seen by $Tr_3$, then the value of $R_8$ is scaled to preserve the same closed-loop gain, and this reduces the voltage drops caused by input transistor base currents.

My previous amplifier designs assumed that a ±50 mV output d.c. offset is acceptable. This allowed d.c. trimming, offset servos, etc. to be greatfully dispensed with. However, it is not in my nature to leave well enough alone, and it could be argued that ±50 mV is on the high side for a top-flight amplifier. For this reason, I have reduced this range as much as possible without resorting to a servo; the required changes were already made when impedance of the feedback network was reduced to minimise Johnson noise. There were details on this in Part I of this article.

With the usual range of component values, the d.c. offset is determined not so much by input transistor $V_{be}$ mismatch, which tends to be only 5 mV or so, but more by a second mechanism – imbalance in beta. This causes imbalance of base currents, $I_b$, drawn thorough input bias resistor $R_1$ and feedback resistor $R_8$. Cancellation of the voltage-drops across these components is therefore compromised.

A third source of d.c. offset is non-ideal matching of input degeneration resistors $R_{2,3}$. Here they are 100 Ω, with 300 mV dropped across each, so two 1% components at opposite ends of their tolerance bands could give a maximum offset of 6 mV. In practice, it is unlikely that the error from this source will exceed 2 mV.

There are several ways to reduce d.c. offset. Firstly, a class-a amplifier with a single output pair must be run from modest ht rails, so the requirement for high-$V_{ce}$ input transistors is relaxed. This allows higher beta devices to

be used, directly reducing $I_b$. The *2SA970* devices used in this design have a beta range of 350 to 700, compared with 100 or less for *MPSA06/56*. Note the pinout is *not* the same.

In chapter 26, we reduced the impedance of the feedback network by a factor of 4.5, and the offset component due to $I_b$ imbalance is reduced by the same ratio. We might therefore hope to keep the d.c. output offset for the improved amplifier to within ±15 mV without trimming or servos. Using high-beta input devices, the $I_b$ errors did not exceed ±15 mV for ten sample pairs – *not* all from the same batch – and only three pairs exceeded ±10 mV. Errors in $I_b$ are now reduced to the same order of magnitude as $V_{be}$ mismatches, and so no great improvement can be expected from further reduction of circuit resistances. Drift over time was measured at less than 1 mV, and this seems to be entirely a function of temperature equality in the input pair.

Figure 1 shows the ideal d.c. conditions in a perfectly-balanced input stage, assuming a β of 400, compared with a set of real voltages and currents

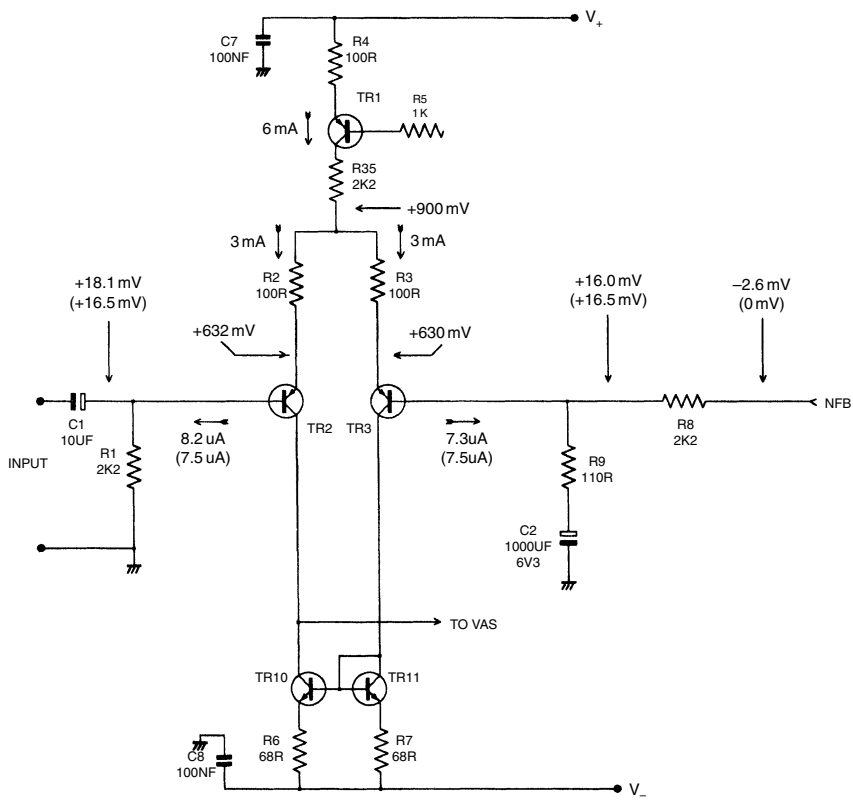



**Figure 1** *A close look at input stage balance. Circuit conditions shown here are a real example. Ideal conditions for* $\beta = 400$ *are shown in brackets. All voltages measured to ground.*

from the prototype amplifier. In the latter case, there is a typical partial cancellation of offsets from the three different mechanisms, resulting in a creditable output offset of −2.6 mV.

### *Biasing for three modes*

Figure 2 shows a simplified rendering of the Trimodal biasing system; the full version appears in Figure 3. The voltage between points A and B is determined by one of which can be in command at a time. Since both are basically shunt voltage regulators sitting between A and B, the result is that the lowest voltage wins. The novel Class-A current-controller introduced in the original article[1] is used here adapted for 0.1 Ω emitter resistors, mainly by reducing the reference voltage to 300 mV, which gives a quiescent current ($I_q$) of 1.5 A when established across the total emitter resistance of 0.2 Ω.

In parallel with the current-controller is the $V_{be}$ multiplier $Tr_{13}$. In Class-B mode, the current-controller is disabled, and critical biasing for minimal crossover distortion is provided in the usual way by adjusting preset $Pr_1$ to

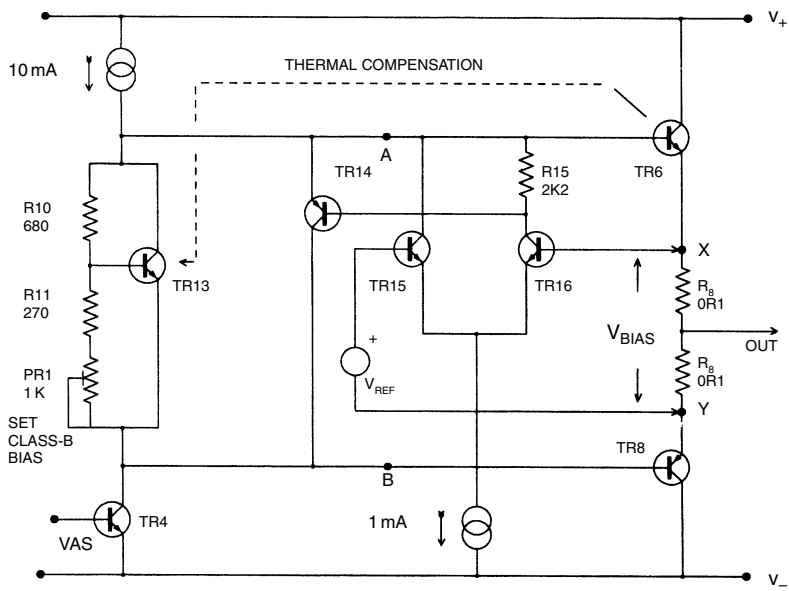



**Figure 2** *Simplified current-controller in action, showing typical d.c. voltages in class-A. Points A, B, X and Y are the same as in the original class-A article. The grey panel on the left is the* $V_{be}$ *multiplier, Class-B biasing and Class-A safety circuit. Panel in the middle is the Class-A current regulator. Voltage over points A, B is* 1.5 V *while over X, Y, i.e.* $V_{bias}$*, there is* 300 mV.

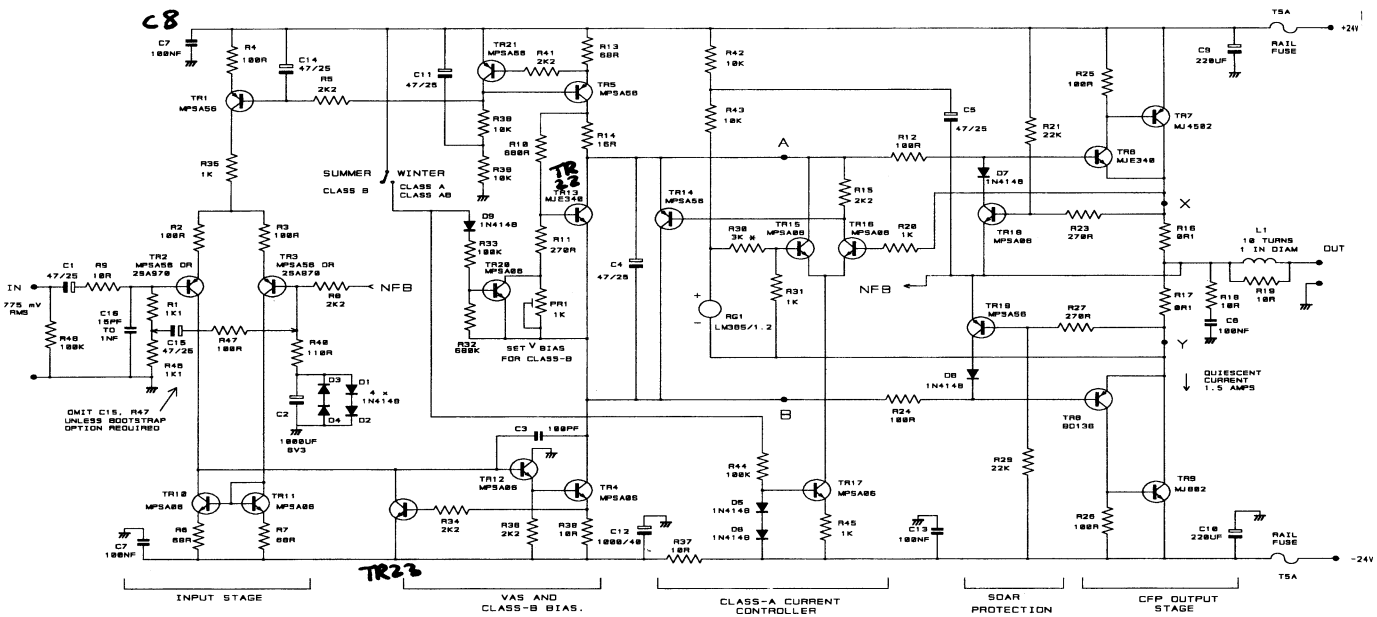



**Figure 3** *Complete circuit diagram of class-A amplifier, including the optional boot-strapping components,* $R_{47}$ *and* $C_{15}$.

set the voltage across $Tr_{13}$. In Class-A/AB mode, the voltage $Tr_{13}$ attempts to establish is increased (by shorting out $Pr_1$) to a value greater than that required for Class-A. The current-controller therefore takes charge of the voltage between X and Y, and unless it fails $Tr_{13}$ does not conduct. Points A B X Y are the same circuit nodes as in reference 1.

### *Class A/AB mode*

In Class-A/AB mode, the current-controller, comprising $Tr_{14,15,16}$ in Figure 2, is active and $Tr_{13}$ is off, as $Tr_{20}$ has shorted out $Pr_1$. Transistors $Tr_{15,16}$ form a simple differential amplifier that compares the reference voltage across $R_{31}$ with the $V_{bias}$ voltage across output emitter resistors $R_{16}$ and $R_{17}$; as explained in Ref. 1, for Class-A this voltage remains constant despite delivery of current into the load. If the voltage across $R_{16,17}$ tends to rise, then $Tr_{16}$ conducts more, turning $Tr_{14}$ more on and reducing the voltage between A and B. $Tr_{14,15,16}$ all move up and down with the amplifier output, and so a tail current-source $Tr_{17}$ is used.

I am aware that the current-controller is more complex than the simple $V_{be}$ multiplier used in most Class-B designs. There is an obvious risk that an assembly error could cause a massive current that would prompt the output devices to lay down their lives to save the rail fuses. The tail-source $Tr_{17}$ is particularly vulnerable because any fault that extinguishes the tail current removes the drive to $Tr_{14}$, the controller is disabled, and the current in the output stage will be very large. In Figure 2 the $V_{be}$-multiplier $Tr_{13}$ acts as a safety-circuit which limits $V_{bias}$ to about 600 mV rather than the normal 300 mV, even if the current-controller is completely non-functional and $Tr_{14}$ fully off. This gives a 'quiescent' of 3A, and I can testify this is a survivable experience for the output devices in the short-term; however they may eventually fail from overheating if the condition is allowed to persist.

There are important points about the current-controller. The entire tail-current for the error-amplifier, determined by $Tr_{17}$, is syphoned off from the voltage amplifier stage current source $Tr_5$. This must be taken into account when ensuring that the upper output half gets enough drive current.

There must be enough tail current available to turn on $Tr_{14}$, remembering that most of $Tr_{16}$ collector-current flows through $R_{15}$, to keep the pair roughly balanced. If you feel moved to alter the voltage-amplifier stage current, remember also that the base current for driver $Tr_6$ is higher in Class-A than Class-B, so the positive slew-rate is slightly reduced in going from Class-B to A.

I must admit that the details of the voltage reference were rather glossed over in Ref. 1, because space was running out fast. The original amplifier

shown last month used a National *LM385/1.2*, its output voltage fixed at 1.223 V nominal; this was reduced to approx 0.6 V by a 1 k Ω/1 k Ω divider.

The circuit also worked well with $V_{ref}$ provided by a silicon diode, 0.6 V being an appropriate bias voltage drop across two 0.22 Ω output emitter resistors. This is simple, and retains the immunity of $I_q$ to heatsink and output device temperatures, but it does sacrifice the total immunity to ambient temperature that a band-gap reference gives.

The *LM385/1.2* is the lowest voltage band-gap reference commonly available; however, the voltages shown in Figure 2 reveal a difficulty with the new lower $V_{bias}$ value and the complementary feedback pair stage; points A and Y are now only 960 mV apart, which does not give the reference room to work in if powered from node A, as in the original circuit.

The solution is to power the reference from the positive rail, via $R_{42,43}$. The midpoint of these two resistors is boot-strapped from the amplifier output rail by $C_5$, keeping the voltage across $R_{43}$ effectively constant. Alternatively, a current-source could be used, but this might reduce positive headroom. Since there is no longer a strict upper limit on the reference voltage, a more easily obtainable 2.56 V device could be used providing $R_{30}$ is suitably increased to 7 kΩ to maintain $V_{ref}$ at 300 mV across $R_{31}$.

In practice, stability of $I_q$ is very good, staying within 1% for long periods. The most obvious limitation on stability is differential heating of $Tr_{15,16}$ due to the main heatsink. Transistor $Tr_{14}$ should also be sited with this in mind, as heating it will increase its beta and slightly imbalance $Tr_{15,16.}$

## Class-B mode

In Class-B mode, the current-controller is disabled, by turning off tail-source $Tr_{17}$ so $Tr_{14}$ is firmly off, and critical biasing for minimal crossover distortion is provided as usual by $V_{be}$-multiplier $Tr_{13}$. With 0.1 Ω emitter resistors $V_{bias}$ (between X and Y) is approx 10 mV. I would emphasise that in Class-B this design, if constructed correctly, will be as 'blameless' as a purpose-built Class-B amplifier. No compromises have been made in adding the mode-switching.

As in the previous Class-B design, the addition of $R_{14}$ to the $V_{be}$-multiplier compensates against drift of the voltage amplifier stage current-source $Tr_5$. To make an old but much-neglected point, the preset potentiometer should always be in the bottom arm of the $V_{be}$ divider $R_{10,11}$ because when presets fail it is usually by the wiper going open; in the bottom arm this gives minimum bias voltage, but in the upper arm it would give maximum.

In Class-B, temperature compensation for changes in driver dissipation remains vital. Thermal runaway with the complementary feedback pair is most unlikely, but accurate quiescent setting is the only away to minimise cross-over distortion. $Tr_{13}$ is therefore mounted on the same small heatsink

as driver $Tr_6$. This is often called thermal feedback, but it is no such thing as $Tr_{13}$ in no way controls the temperature of $Tr_6$; 'thermal feedforward' would be a more accurate term.

### *Switching modes*

The dual nature of the biasing system means Class-A/Class-B switching is easily implemented, as in Figure 3. A Class-A amplifier is an uneasy companion in hot weather, and so I was unable to resist the temptation to sub-title the mode switch 'Summer/Winter', by analogy with a car air intake.

Switchover is d.c.-controlled, as it is not desirable to have more signal than necessary running around inside the box, possibly compromising inter-channel crosstalk. In Class-A/AB mode, $S_1$ is closed, so $Tr_{17}$ is biased normally by $D_{5,6}$, and $Tr_{20}$ is held on via $R_{33}$, shorting out present $Pr_1$ and setting $Tr_{13}$ to safety mode, maintaining a maximum $V_{bias}$ limit of 600 mV. For Class-B, $S_1$ is opened, turning off $Tr_{17}$ and therefore $Tr_{15,16}$ and $Tr_{14}$. Transistor $Tr_{20}$ also ceases to conduct, protected against reverse-bias by $D_9$, and reduces the voltage set by $Tr_{13}$ to a suitable level for Class-B. The two control pins of a stereo amplifier can be connected together, and the switching performed with a single-pole switch, without interaction or increased crosstalk.

Mode-switching affects the current flowing in the output devices, but the output voltage is controlled by the global feedback loop, and switching is completely silent in operation. The mode is switchable while the amplifier is handling audio, allowing some interesting 'A/B' listening tests.

It may be questioned why it is necessary to explicitly disable the current-controller in Class-B; $Tr_{13}$ is establishing a lower voltage than the current-controller which latter subsystem will therefore turn $Tr_{14}$ off as it strives futilely to increase $V_{bias}$. This is true for 8 Ω loads, but 4 Ω impedances increase the currents flowing in $R_{16,17}$ so they are transiently greater than the Class-A $I_q$, and the controller will therefore intermittently take control in an attempt to reduce the average current to 1.5A. Disabling the controller by turning off $Tr_{17}$ via $R_{44}$ prevents this.

## No warm up

Audio magazines often state that semiconductor amplifiers sound better after hours of warm-up. If this is true – in most cased it almost certainly isn't – the admission represents truly spectacular design incompetence. Accusations of this type are applied with particular venom to class-A designs, because it is obvious that the large heat sinks required take time to reach final temperature. So it is important to record that in

class-A operation this design stabilises its electrical operating conditions in less than a second, giving the full intended performance.

No 'warm-up time' beyond this is required.

Obviously the heat sinks take time to reach thermal equilibrium. But as already described, measures have been taken to ensure that component temperature has no significant effect on operating conditions or performance.

## Supplying power

Regulated supplies are quite unnecessary, and are virtually certain to do more harm than a good unregulated power supply (Figure 4).

The supply must be designed for continuous operation at maximum current, so the bridge rectifier should be properly heat-sunk, and careful consideration given to the ripplecurrent ratings of the reservoirs. This is one reason why reservoir capacitance has been doubled to 20,000 μF per rail: the ripple voltage is halved, improving voltage efficiency as it is the ripple troughs that determine clipping onset. But the ripple current, although unchanged in total value, is now split between two components. (The capacitance was not increased to reduce ripple injection. This is dealt with far more efficiently and economically by making amplifier psrr high.[3])

Do not omit the secondary fuses. Even in these modern times rectifiers do fail, and transformers are horribly expensive.

## Test mode

If the Class-A controller is enabled, but preset $Pr_1$ is left in circuit, (e.g. by shorting $Tr_{20}$ base-emitter) we have a test mode which allows suitably cautious testing; current $I_q$ is zero with the preset fully down, as $Tr_{13}$ overrides the current-controller, but increases steadily as $Pr_1$ is advanced, until it suddenly locks at the desired quiescent current. If the current-controller is faulty then $I_q$ continues to increase to the defined maximum of 3A.

## Thermal design

Class-A amplifiers are hot almost by definition, and careful thermal design is needed if they are to be reliable, and not take the varnish off the Sheraton. Since the internal dissipation of the amplifier is maximal with

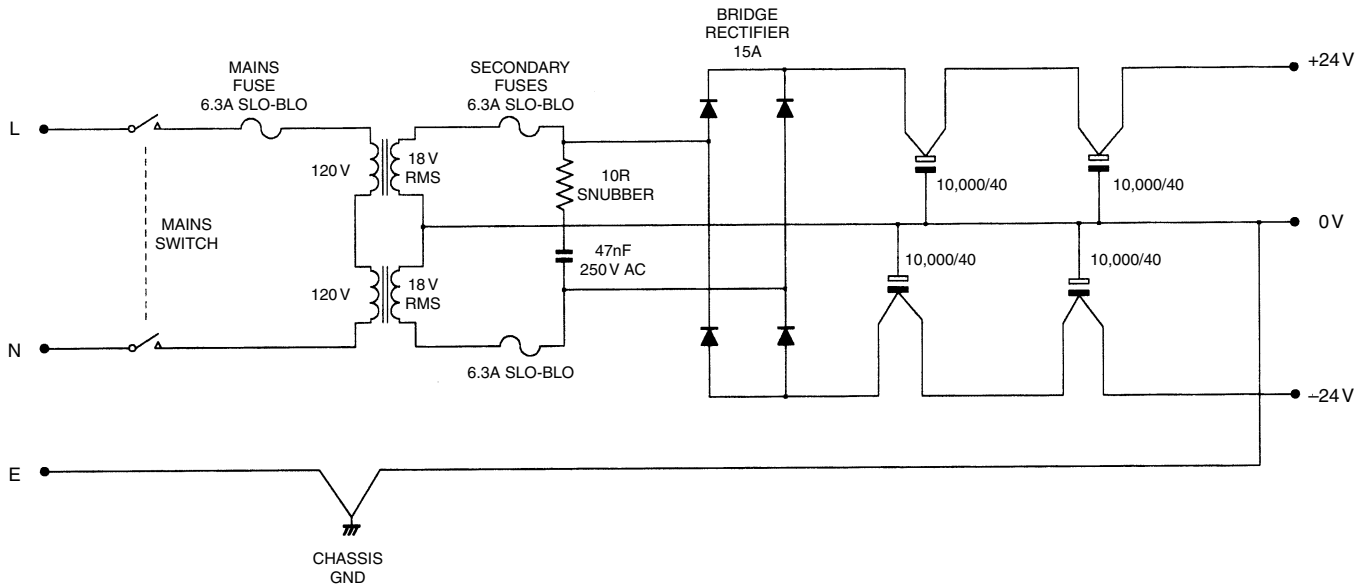



**Figure 4** *Power supply circuit diagram.*

**Table 1** *Power capability of the trimodal power amplifier*

| | *W* | *W* | *W* | *Distortion* |
|---|---|---|---|---|
| Load resistance (Ω) | 8 | 6 | 4 | |
| Class A | 20 | 27 | 15 | low |
| Class AB | n/a | n/a | 39 | high |
| Class B | 21 | 28 | 39 | medium |

no signal, simply turning on the prototype and leaving it to idle for several hours will give an excellent idea of worst-case component temperatures. In Class-B the power dissipation is very programme-dependant, and estimates of actual device temperatures in realistic use are notoriously difficult.

Table 1 shows the output power available in the various modes, with typical transformer regulation, etc; the output mode diagram in Part 1, Figure 1, showed exactly how the amplifier changes mode from A to AB with decreasing load resistance. Remember that in this context 'high distortion' means 0.002% at 1 kHz. This diagram was produced in the analysis section of *PSpice* simply by typing in equations, and without actually simulating anything at all.

The most important thermal decision is the size of the heatsink; it is going to be expensive, so there is a powerful incentive to make it no bigger than necessary. I have ruled out fan cooling as it tends to make concern for ultra-low electrical noise look rather foolish; let us rather spend the cost of the fan on extra cooling fins and convect in ghostly silence. The exact thermal design calculations are simple but tedious, with many parameters to enter; the perfect job for a spreadsheet. The final answer is the margin between the predicted junction temperatures and the rated maximum.

Once power output and impedance range is decided, the heatsink thermal resistance to ambient is the main variable to manipulate; and this is a compromise between coolness and cost, for high junction temperatures always reduce semi-conductor reliability, Table 2.

**Table 2** *Temperature considerations*

| | *Thermal resist* °C/W | *heat flow* W | *temp rise* °C | *temp* °C |
|---|---|---|---|---|
| Juncn to to 3 case | 0.7 | 36 | 25 | 100 junction |
| Case to sink | 0.23 | 36 | 8 | 75 TO3 case |
| Sink to air | 0.65 | 72 | 47 | 67 heatsink |
| Total | | | 80 | 20 ambient |

This shows that the transistor junctions will be 80 °C above ambient, i.e. at around 100 °C; the rated junction maximum is 200 °C, but it isn't wise to get anywhere close to this very real limit. Note the *Case-Sink* thermal washers are made from high-efficiency material. Standard versions have a slightly higher thermal resistance.

The heatsinks used in the prototype had a thermal resistance of 0.65 °C/W per channel. This is a substantial piece of metal, and is expensive.

## The complete circuit

The complete Class-A amplifier is shown in Figure 3, complete with optional input bootstrapping but omitting any balanced-line input amplifier or gain control. The circuitry may look a little complex at first, but we have only added four low-cost transistors to realise a high-accuracy Class-A quiescent controller, and one more to implement mode-switching. Since the biasing system has been described above, only the remaining amplifier subsystems are dealt with here.

The input stage follows my design methodology in running at a high tail current to maximise transconductance, and then linearizing it by adding input degeneration resistors $R_{2,3}$. These reduce the final transconductance to a suitable level. Current-mirror $Tr_{10,11}$ forces the collector currents of the two input devices $Tr_{2,3}$ to be equal, balancing the input stage to prevent the generation of second-harmonic distortion. The mirror is degenerated by $R_{6,7}$ to eliminate the effects of $V_{be}$ mismatches in $Tr_{10,11}$.

With some misgivings I added the input network $R_9$, $C_{15}$, which is definitely *not* intended to define the system bandwidth, unless fed from a buffer stage; with practical values the h.f. roll off could vary widely with the source impedance driving the amplifier. It is intended rather to give the possibility of dealing with rf interference without having to cut tracks. Resistor $R_9$ could be increased for bandwidth definition if the source impedance is known, fixed, and taken into account when choosing $R_9$; bear in mind that any value over 47 Ω will measurably degrade the noise performance. The values given roll off above 150 MHz to keep out uhf.

As a result of insights gained while studying the slewing behaviour of the generic/Lin configuration, I have increased the input-stage tail current from 4 to 6 mA, and increased the voltage amplifier stage standing current from 6 to 10 mA over the original circuit. This increases the maximum positive and negative slew rates from the basic +21, −48 V/μs of reference 4 to +37, −52 V/μs; as described elsewhere[2] this amplifier architecture is always assymetrical in slew rate. One reason is feedthrough in the voltage amplifier current source; in the original circuit an unexpected slew rate limit was set by fast edges coupling through the current source c-b capacitance to reduce the bias voltage during positive slewing. This effect is minimised here by using the

negative-feedback type of current source bias generator, with voltage amplifier collector current chosen as the controlled variable.

Transistor $Tr_{21}$ senses the voltage across $R_{13}$, and if it attempts to exceed $V_{be}$, turns on further to pull up the bases of $Tr_1$ and $Tr_5$. Capacitor $C_{11}$ filters the d.c. supply to this circuit and prevents ripple injection from the positive rail. Capacitor $C_{14}$, with $R_5$, provides decoupling. Increasing input tail-current also mildly improves input-stage linearity, as it raises the basic transistor $g_m$ and allows $R_{2,3}$ to apply more local feedback.

The voltage amplifier stage is linearised by beta-enhancing stage $Tr_{12}$, which increases the amount of local feedback through Miller dominant-pole capacitor $C_3$, often referred to as $C_{dom}$. Resistor $R_{36}$ has been increased to 2.2 kΩ to minimise power dissipation, as there seems to be no significant effect on linearity or slewing. Do not, however, attempt to omit it altogether, or linearity *will* be affected and slewing much compromised.

As described in Ref. 3, the simplest way to prevent ripple from entering the voltage amplifier via the negative rail is old-fashioned *RC* decoupling, with a small *R* and a big *C*. We have some 200 mV in hand (Chapter 26) in the negative direction, compared with the positive, and expending this as the voltage-drop through the *RC* decoupling will give symmetrical clipping. $R_{37}$ and $C_{12}$ perform this function; the low rail voltages in this design allow the 1000 µF capacitor $C_{12}$ to be a fairly compact component.

The output stage is of the complementary feedback pair (CFP) type. As described in Chapter 26, this gives the best linearity and quiescent stability, due to the two local negative feedback loops around driver and output device. Quiescent stability is particularly important with $R_{16,17}$ as low as 0.1 Ω, and this low value would probably be rather dicey in a double emitter-follower output stage.

Voltage efficiency of the complementary feedback pair is also higher than the emitter follower version. Resistor $R_{25,26}$ define a suitable quiescent collector current for the drivers $Tr_{6,8}$, and pull charge carriers from the output device bases when they are turning off. The lower driver is now a *BD136*; this has a higher $f_T$ than the *MJE350*, and seems to be more immune to odd parasitics at negative clipping.

The new lower values for the output emitter resistors $R_{16,17}$ halve the distortion in Class-AB. This is equally effective when in Class-A with too low a load impedance, or in Class-B but with $I_q$ maladjusted too high. It is now true in the latter case that too much $I_q$ really is better than too little – but not much better, and AB still comes a poor third in linearity to Classes A and B.

Safe operating area protection is given by the networks around $Tr_{18,19}$. This is a single-slope safe operating area system that is simpler than two-slope safe area, and therefore somewhat less efficient in terms of snuggling the limiting characteristic up to the true safe operating area of the output transistor. However, in this application, with low rail voltages,

## An adaptive trimodal design?

One interesting extension of the ideas presented here is the adaptive trimodal amplifier. This would switch into class-B on detecting device or heat-sink over-temperature, and would be a unique example of an amplifier that changed mode to suit the operating conditions.

Thermal protection would need to be latching as flipping from class-A to class-B every few minutes would subject the output devices to unnecessary thermal cycling.

maximum utilisation of the transistor safe area is not really an issue; the important thing is to observe maximum junction temperatures in the A/AB mode.

The global negative-feedback factor is 32 dB at 20 kHz, and this should give a good margin of safety against Nyquist-type oscillation. Global negative feedback increases at 6 dB/octave with decreasing frequency to a plateau of around 64 dB, the corner being at a rather ill-defined 300 Hz; this is then maintained down to 10 Hz. It is fortunate that magnitude and frequency here are non-critical, as they depend on transistor beta and other doubtful parameters.

## Performance

The performance of a properly-designed Class-A amplifier challenges the ability of even the Audio Precision measurement system. To give some perspective on this, Figure 5 shows the distortion of the AP oscillator

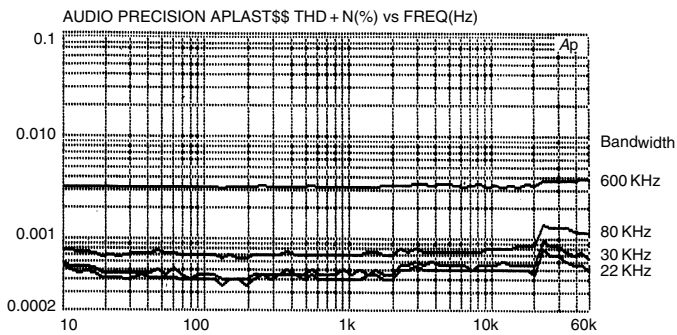



**Figure 5** *Distortion plot of the Audio Precision oscillator/analyser combination alone, for measurement bandwidths of* 500, 80, 30 *and* 22 kHz. *The saw-teeth below* 1 kHz *are artifacts. The residual appears to be pure noise.*

driving the analyser section directly for various bandwidths. There appear to be internal mode changes at 2 kHz and 20 kHz, causing step increases in oscillator distortion content; these are just visible in the THD plots for Class-A mode.

Figure 6 shows Class-B distortion for 20 W into 8 and 4 Ω, while Figure 7 shows the same in Class-A/AB.

I would like to acknowledge the invaluable help and encouragement of Gareth Connor. Credit goes to him for the tricky task of PCB layout – and not me, as previous adverts have implied.

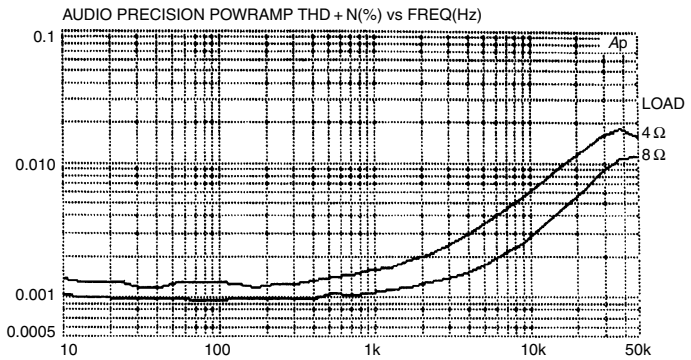



**Figure 6** *Distortion in class-B (summer) mode. Distortion into* 4 Ω *is always worse. Power was* 20 W *in* 8 Ω *and* 40 W *in* 4 Ω*, bandwidth* 80 kHz.

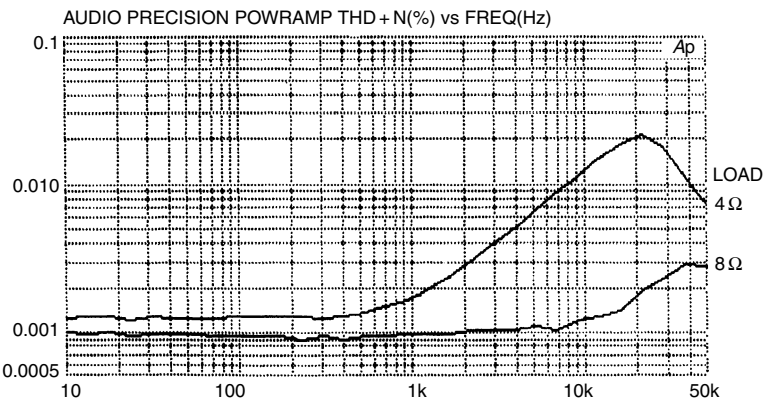



**Figure 7** *Distortion in class-A/AB (winter) mode, same power and bandwidth. The amplifier is in AB mode for the* 4 Ω *case, and so distortion is higher than for class-B*Ω*. At* 80 kHz *bandwidth, the class-A plot below* 10 kHz *merely shows the noise floor.*

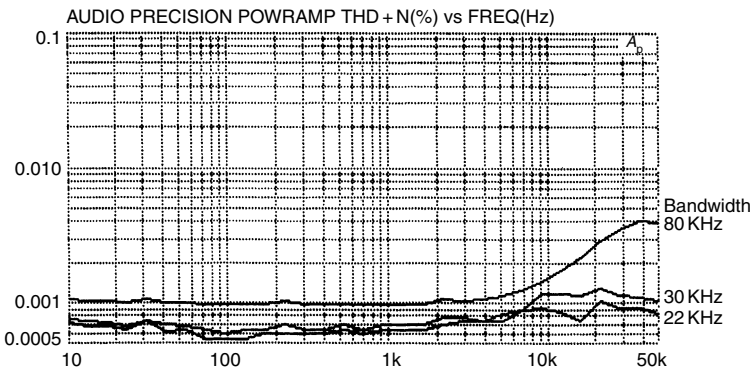



**Figure 8** *Distortion in class-A only* (20 W/8 Ω) *for varying measurement bandwidths. The lower bandwidths ignore h.f. distortion, but give a much clearer view of the excellent linearity below* 10 kHz.

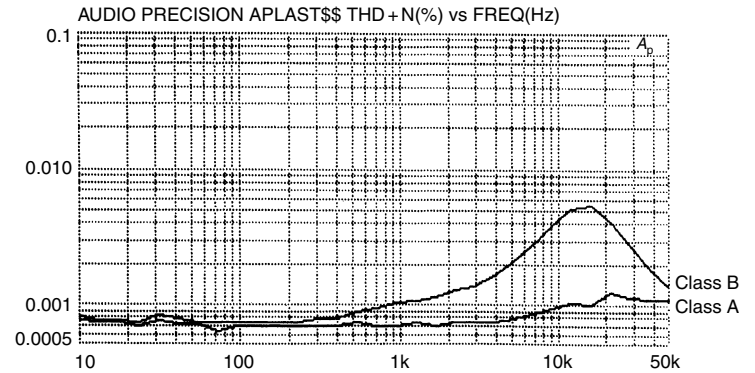



**Figure 9** *Direct comparison of classes A and B* (20 W/8 Ω) *at* 30 kHz *bandwidth. The h.f. rise for B is due to the inability of negative feedback that falls with frequency to linearise the high-order crossover distortion in the output stage.*

## References

1. Self, D. 'Distortion in power amplifiers; part 8', *Electronics World & Wireless World*, March 94, p 225.
2. Self, D. 'High speed audio power', *Electronics World & Wireless World*, September 1994, p 760.
3. Self, D. 'Off the rails', *Electronics World & Wireless World*, March 1995, p 201.
4. Self, D. 'Distortion in power amplifiers; Part 7', *Electronics World & Wireless World*, February 1994, p 139.

## *Erratum*

Regrettably, a couple of errors crept into the original article on Class-A.[1] On page 229, second column: '$Tr_{15,16}$ then compares the reference voltage with that at point Y' should read 'at point X'. On page 229, third column: 'This comes to the same thing as maintaining a constant $V_{bias}$ across $Tr_5$' should read 'across $Tr_{13}$'. This is nobody's fault but mine, and I humbly apologise as it cannot have made understanding the current-controller action any easier. (Author's note: these corrections have been made in this edition.)

# 28 Load-invariant audio power

*January 1997*

> It has been my experience that power amplifiers always, without exception, give more distortion with heavier loading i.e., with a lower load impedance. This, in BJT amplifiers at least, is due to an extra distortion mechanism that becomes more and more important as the load impedance falls and the output stage currents increase. Against all intuition, the source of the increased distortion is the drivers, and not the output devices themselves; this was wholly unexpected, and to the best of my knowledge was a new discovery. It was uncovered not so much by mathematics or theory as by some careful SPICE analysis, later backed up by measurements, and was a copybook example of just how very powerful and useful circuit simulation can be in the right circumstances.
>
> To see if this extra distortion could be eliminated, or at any rate reduced with respect to its 'normal' levels, I conducted a little research program, and this chapter was the result. It was not possible to make the amplifier totally Load-Invariant, i.e., with the same THD at 4 Ω as 8 Ω, but I think I got pretty close.

My investigations into power amplifiers have so far largely concentrated on 8 Ω resistive loading. This is open to criticism, as loudspeaker impedance dips to 4 Ω or less are not uncommon. Solid-state amplifiers always give more distortion with heavier loading, without exception so far as I am aware.

While it would be highly desirable from the amplifier designer's point of view for the loudspeaker designer to strive for a reasonably flat impedance, it has to be accepted that electronic problems are much easier to solve than electromechanical ones. It follows that it is reasonable for amplifiers to accommodate themselves to loudspeakers rather than the other way around. Thus an amplifier must be able to cope gracefully with impedance dips to 4 Ω or lower.

Such dips tend to be localised in frequency, so music does not often dwell in them. An amplifier should be capable of driving half the nominal load impedance at almost the full voltage swing, though not necessarily for more than a minute or so.

Contemporary power amplifier ratings tend to be presented in the format '*X* watts into 8 Ω, *Y* watts into 4 Ω' from which we presumably may deduce:

- The amplifier will deliver sustained power into 4 Ω.
- Since 2 Ω loads are not explicitly mentioned, they cannot be driven in a sustained fashion.

It may also be assumed, but with much less certainty, that,

- The amplifier will cope with short-term 2 Ω impedance dips; i.e. half the lowest nominal load quoted.
- The overload protection – if it exists at all – activates below 2 Ω. Note that no minimum load impedance is specified.

## Output loading and distortion

A 'Blameless' Class-B power amplifier is one wherein all the distortion mechanisms shown in Table 1 have been eliminated or reduced to below the noise floor, except for the intractable Distortion 3 in its three subcategories. I have produced a slim monograph which describes the philosophy and practicalities of this in greater detail than *EW* articles permit.[1]

A Blameless design gives a distortion performance into 8 Ω that depends very little on variable transistor characteristics such as beta. This is because

**Table 1** *Characteristics of distortion mechanisms.*

| *No.* | *Mechanism* | *Category* | *Component sensitive?* |
|---|---|---|---|
| 1 | Input $V_{in}/I_{out}$ nonlinearity | Inherent | No |
| 2 | VAS $I_{in}/V_{out}$ nonlinearity | Inherent | Yes? |
| 3 | Output stage distortions: | | |
| | a) Large-signal nonlinearity | Inherent | Yes |
| | b) Crossover distortion | Inherent | No? |
| | c) Switch-off distortion | Inherent | Yes |
| 4 | Non-linear voltage-amplifier stage loading | Inherent | Yes |
| 5 | Rail decouple grounding | Topological | No |
| 6 | Rail current induction | Topological | No |
| 7 | Error in negative-feedback take-off-point | Topological | No |
| 8 | Feedback cap distortion | Inherent | Yes |

at this load impedance the output stage nonlinearity is almost all crossover distortion, which is primarily a voltage-domain effect.

Note that for optimal crossover behaviour the quantity to be set is $V_q$, the voltage across the two output emitter resistors $R_e$, and the actual value of the resulting $I_q$ is incidental.[2] Mercifully, in Class-B the same $V_q$ remains optimal whatever the load impedance; if it did not the extra complications would be serious.

As the load impedance of a Blameless Class-B amplifier is decreased from infinite to 4 Ω, distortion increases in an intriguing manner. Unloaded, the THD is not much greater than that from the Audio Precision test oscillator, but with loading crossover distortion increases steadily, Figure 1.

When the load impedance falls below about 8 Ω, a new distortion begins to appear, overlaid on the existing crossover nonlinearities. It is low-order, and essentially third-harmonic. In Figure 2 the upper 4 Ω THD trace is consistently twice that for 8 Ω, once it clears the noise floor.

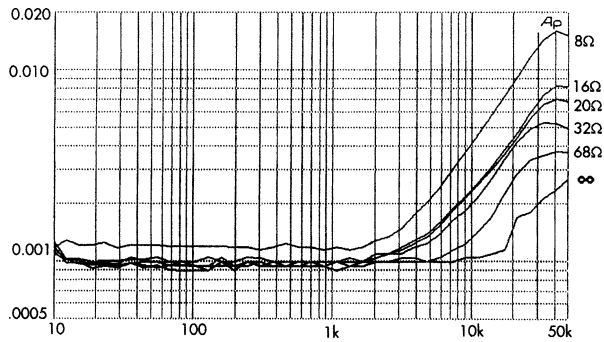



**Figure 1** *Crossover distortion from a Blameless amplifier increases as load resistance falls to* 8 Ω*. All plots at* 80 kHz *bandwidth.*

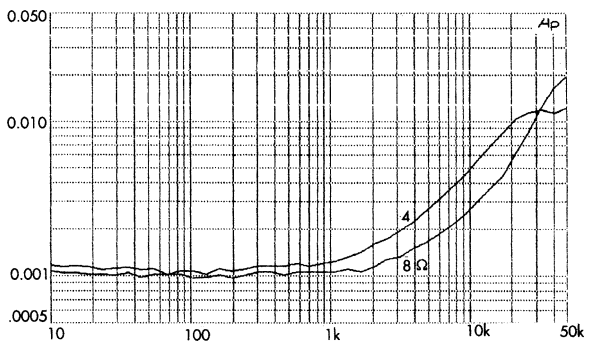



**Figure 2** *Upper trace shows distortion increase due to large-signal nonlinearity as load goes from* 8 *to* 4 Ω*. Blameless amplifier at* 25 W/8 Ω*.*

In Chapter 20, I labelled this as Distortion 3a, or large signal nonlinearity. The word 'large' refers to currents rather than voltages. Unlike crossover distortion 3b, the amount of LSN produced is significantly dependant on device characteristics.[3] The distortion residual is essentially third-order due to the symmetric and compressive nature of the output stage gain characteristic, but its appearance on a scope can be complicated by different amounts of nonlinearity in the upper and lower output stage halves.

Large signal nonlinearity occurs in both emitter-follower and complementary feedback pair output configurations; this chapter concentrates on the complementary feedback pair, as in Figure 3. Incremental gain of a simulated complementary feedback pair output stage for 8 and 4 Ω is shown in Figure 4; the lower 4 Ω trace has greater downward curvature, i.e. a greater fall off of gain with increasing current. Simulated emitter follower behaviour is similar.

As it happens, an 8 Ω nominal impedance is a pretty good match for standard power bipolar junction transistors, though 16 Ω might be better for minimising large-signal nonlinearity – loudspeaker technology permitting. It is presumably coincidental that the 8 Ω nominal impedance corresponds approximately with the heaviest load that can be driven without large signal nonlinearity appearing.

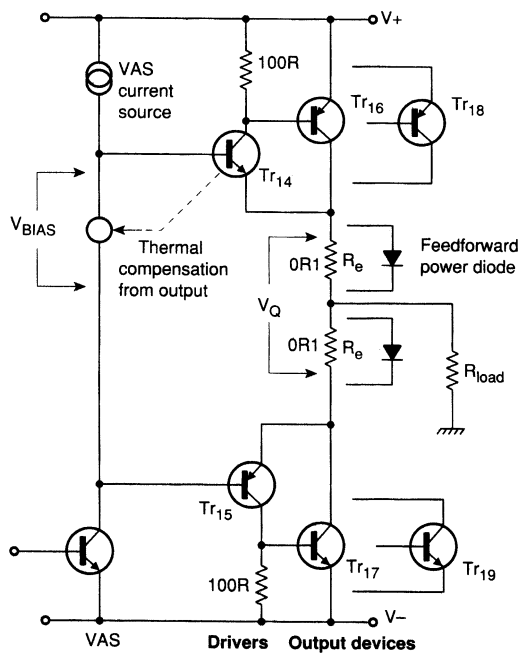



**Figure 3** *Complementary feedback pair output stage, showing how extra devices are added in parallel, and where feedforward diodes would be fitted.*

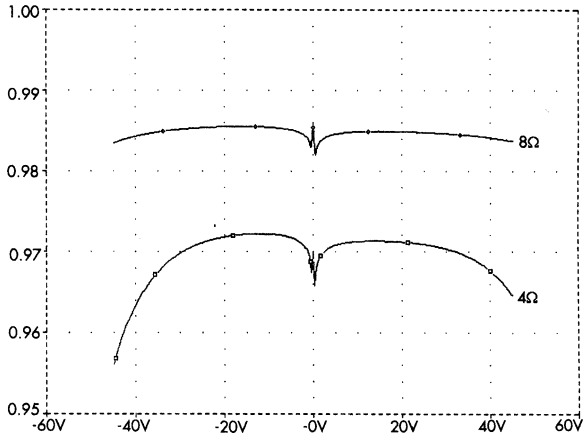



**Figure 4** *Incremental gain of a standard complementary feedback pair output stage. The* 4 Ω *trace droops much more as the gain falls off at higher currents. (PSpice)*

Since large signal nonlinearity is an extra distortion component laid on top of others, and usually dominating them in amplitude, it is obviously simplest to minimise the 8 Ω distortion first, so that 4 Ω effects can be seen more or less in isolation when they appear.

The typical result of 4 Ω amplifier loading was shown in Figure 2, for the relatively modern *MJ15024/25* complementary pair from Motorola. Figure 5 shows the same for one of the oldest silicon complementary pairs, the *2N3055/2955*, unfortunately on a slightly different frequency scale. The 8 Ω distortion is similar for the different devices, but the 4 Ω THD is 3.0 times worse for the venerable *2N3055/2955*. Such is progress.

Such experiments with different output devices throw useful light on the Blameless concept. From various types tried so far it can be said that

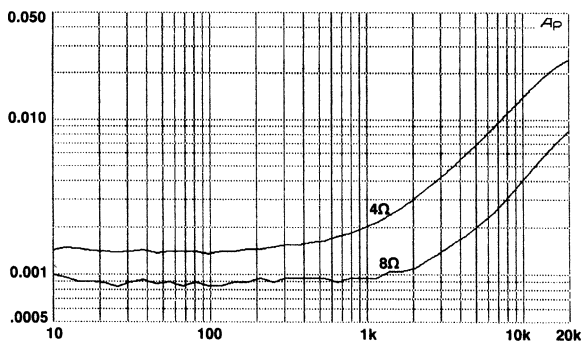



**Figure 5** *Distortion with* 4 Ω *load is* 3× *greater than* 8 Ω *for 2N3055/2955 output devices. Compare Figure 2.*

Blameless performance, independent of output device type, should not exceed 0.001% at 1 kHz and 0.006% at 10 kHz, into 8 Ω. All the components existed to build sub-0.001% THD amplifiers in mid-1969 – if only we had known how to do it.

Low-impedance loads have other implications beyond worsening the THD. The requirements for long-term 4 Ω operation are severe, demanding significantly more heatsinking and power supply capacity if reliability is to be maintained.

For economic reasons the peak–average ratio of music is usually fully exploited, though this can cause real problems on extended tests, such as the FTC 40%-power-for-an-hour preconditioning procedure.

The main subject of this article is the extra distortion generated in the output stage itself by increased loading, but there are other ways in which the total amplifier distortion may be degraded by the increased currents flowing.

Table 1 shows the main distortion mechanisms in a power amplifier; Distortions 1, 2, and 8 are unaffected by output stage conditions. Distortion 4 might be expected to increase, as the increased loading on the output stage is reflected in increased voltage amplifier stage loading.[4] However, both the beta-enhanced emitter-follower and buffered-cascode methods seem to cope effectively with sub-8 Ω loads.

The greater supply currents drawn could increase the rail ripple, which will worsen Distortion 5 if it exists. But since the supply reservoir capacitance must also be increased to permit greater power delivery, ripple will be reduced again and this tends to cancel out. If the rail ripple does increase, the usual *RC* filtering of bias supplies[5] deals with it effectively, preventing it getting in via the input pair tail, etc.

Distortion 6 may be more difficult to eliminate as the halfwave currents flowing in the output circuitry are twice as large, with no counteracting mechanism. Distortion 7, if present, will be worsened due the increased load currents flowing in the output stage wiring resistances.

Of those mechanisms above, Distortion 4 is inherent in the circuit configuration – though not a problem in practice – while 5, 6, and 7 are topological, in that they depend on the spatial and geometrical arrangements of components and wiring. The latter three can therefore be completely eliminated in both theory and practice. This leaves us with only the large signal nonlinearity component of Distortion 3 to grapple with.

## The load-invariant concept

Ideally, the extra distortion component large signal nonlinearity would not exist. Such an amplifier would give no more distortion into 4 Ω than 8, and I call it 'load-invariant to 4 Ω'. The loading qualification is required

because, as you will see, the lower the impedance, the greater the difficulties in aspiring to load-invariance.

I am assuming that we start out with an amplifier that is Blameless at 8 Ω; it would be logical but pointless to apply the term 'load-invariant' to an ill-conceived amplifier delivering 1% THD into both 8 and 4 Ω.

## Large signal nonlinearity

Large signal nonlinearity is clearly a current-domain effect, dependent on the magnitude of the signal currents flowing in drivers and output devices, as the voltage conditions are unchanged.

A 4 Ω load doubles the output device currents, but this does not in itself generate significant extra distortion. The crucial factor appears to be that the current drawn from the drivers by the output device bases *more* than doubles, due to beta fall-off in the output devices with increasing collector current. It is this *extra* increase of current due to beta-droop that causes almost all the additional distortion.

The exact details of how this works are not completely clear, but seems to be because the 'extra current' due to beta fall-off varies very non linearly with output voltage. It appears that the non linear extra current combines with driver nonlinearity in a particularly pernicious way. Beta-droop is ultimately due to what are called high-level injection effects. These vary with device type, so device characteristics now matter.

As I stated in my original power-amplifier series,[6] there is good simulator evidence that large signal nonlinearity is entirely due to the beta-droop causing extra current to be drawn from the drivers. To recapitulate:

- Simulated output stages built from output devices modified to have no beta-droop (by increasing Spice model parameter IKF) have no large signal nonlinearity. It seems to be specifically the extra current taken due to beta-droop that causes the trouble.
- Simulated output devices driven with zero-impedance voltage sources instead of transistor drivers show no large signal nonlinearity. This shows that such nonlinearity does not occur in the outputs themselves, but in the driver transistors.
- Output stage distortion can be regarded as an error voltage between input and output. The double emitter-follower emitter-follower stage error is driver $V_{be}$ + output $V_{be} + R_e$ drop. A simulated emitter-follower output stage with the usual drivers shows that it is primarily nonlinearity in the driver $V_{be}$ that increases as the load resistance reduces, rather than in the output $V_{be}$. The drop across $R_e$ is essentially linear.

These three results have naturally been rechecked for this article.

Knowing that beta-droop caused by increased output device $I_c$ is at the root of the problem leads to some solutions. Firstly, the per-device $I_c$ can be reduced by using parallel output devices. Alternatively $I_c$ can be left unchanged and output device types selected for the least beta droop.

Feedforward diodes across the emitter resistors sometimes help, but they treat the symptoms – by attempting distortion cancellation – rather the root cause, so it is not surprising this method is much less effective.

## Doubled output devices

The basic philosophy here, indicated above, is that the output devices are doubled even though this is quite unnecessary for handling the power output required.

The fall-off of beta depends on collector current. If two output devices are connected in parallel, the collector current divides in two between them, and beta-droop is much reduced. From the above evidence, I predicted that this ought to reduce large-signal nonlinearity and when measured, indeed it does.

This sort of reality-check must never be neglected when you are using simulations. Figure 6 compares 4 Ω THD at 60 W for single and doubled output devices, showing that doubling reduces distortion by about 1.9 times; well worthwhile. The output transistors were standard power devices, in this case Motorola *MJ15024/15025*.

The *2N3055/2955* complementary pair give a similar halving of large-signal nonlinearity on being doubled, though the initial distortion is three times higher into 4 Ω. Those *2N3055s* with an H suffix are markedly worse than those without.

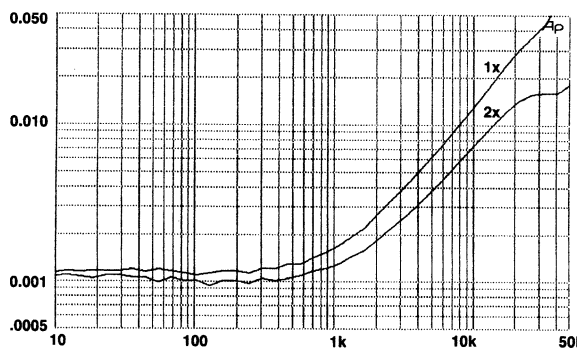



**Figure 6** *Distortion with* 4 Ω *load is reduced by* 1.9× *upon doubling standard MJ15024/15025 output transistors* 30 W/8 Ω.

No current-sharing precautions were taken when doubling the devices, and this lack seemed to have no effect on large-signal nonlinearity reduction. There was no evidence of current-hogging.

Doubling the power devices naturally increases the power output capability, though if this is fully exploited large-signal nonlinearity will tend to rise again, and you are back where you started. It will also be necessary to uprate the power supply and so on. The essence of this technique is to use parallel devices to reduce distortion long before power handling alone compels you to do so.

## Better output devices

The TO3P-packaged *2SC3281* and *2SA1302* complementary pair has a reputation in the hi-fi world for being 'more linear' than the run of transistors. This is the sort of vague claim that arouses the deepest of suspicions, and is comparable with the many assertions of superior linearity in power fets, which is the exact opposite of reality.[7]

In this case however, the kernel of truth is that the *2SC3281* and *2SA1302* show much less beta-droop than average power transistors. These devices were introduced by Toshiba; Motorola versions are *MJL3281A* and *MJL1302A*, also in TO3P. Figure 7 shows beta-droop, for the various devices discussed here, and it is clear that more droop means more large-signal nonlinearity.

The *3281/1302* pair is clearly in a different class from more conventional transistors as regards maintenance of beta with increasing collector

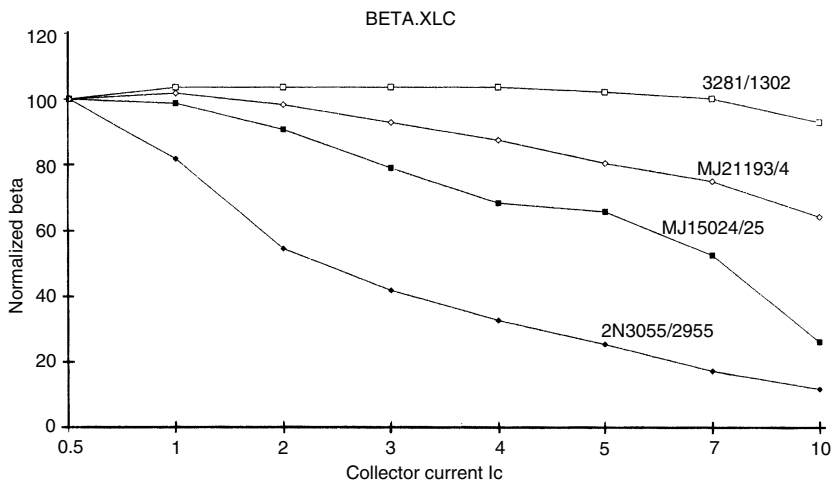



**Figure 7** *Power transistor beta-droop as collector current increases. Beta is normalised to* 100 *at* 0.5 A *based on manufacturers' data sheets.*

current. There seems to be no special name for this class of bipolar junction transistors, so I have called them 'sustained-beta' devices here.

Into 4 and 8 Ω, the THD for single *3281/1302* devices is shown in Figure 8. Distortion is reduced by about 1.4 times compared with the standard devices of Figure 2, over 2–8 kHz. Several pairs of *3281/1302* have been tested and the 4 Ω improvement is consistent and repeatable.

The obvious next step is to combine the two techniques by using double sustained-beta devices. Doubled device results are shown in Figure 9 where the distortion at 80 W/4 Ω (15 kHz) is reduced from 0.009% in Figure 8 to 0.0045% – in other words halved. The 8 and 4 Ω traces are now very close, the 4 Ω THD being only 1.2 times higher than the 8 Ω case.

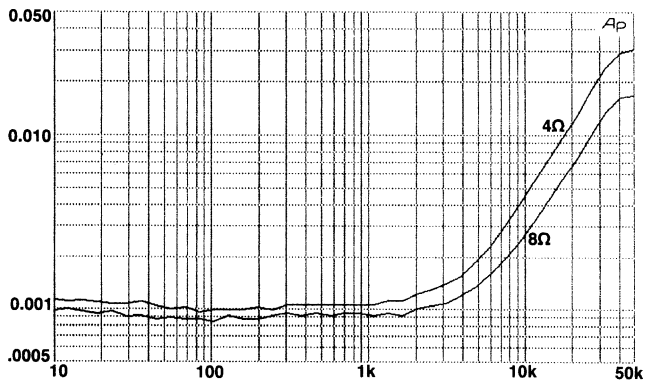



**Figure 8** *Total harmonic distortion at* 40 W/8 Ω *and* 80 W/4 Ω *with single 3281/1302 devices.*

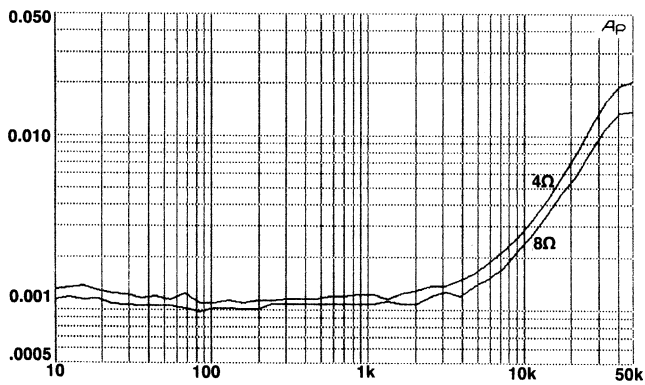



**Figure 9** *At* 40 W/8 Ω *and* 80 W/4 Ω *with doubled 3281/1302 output transistors, the total harmonic distortion looks like this.* 4 Ω *THD has been halved compared with Figure 8.*

Some similar devices exist. Other devices showing less beta-droop than standard are *MJ21193, MJ21194*, in TO3 packaging, and *MJL21193, MJL21194* in TO3P, also from Motorola. These devices show beta-maintenance intermediate between the 'super' *3281/1302* and 'ordinary' *MJ15024/25*, so it seemed likely that they would give less large-signal nonlinearity than ordinary power devices, but more than the *3281/1302*. This prediction was happily fulfilled.

It could be argued that multiplying output transistors is an expensive way to solve a problem. To give this perspective, in a typical stereo power amplifier, including heatsink, metal work and mains transformer, doubling the output devices will only increase the total cost by about 5%.

## Feeding forward

In the Distortion in Power Amplifiers series, the only technique I could offer for improving large-signal nonlinearity was the use of power diodes across 0.22 Ω output emitter resistors.[8] The improvement was only significant for high power into less-than 3 Ω loading, and was of doubtful utility for hifi.

It is now my practice to make output emitter resistors $R_e$ 0.1 Ω, rather than the more usual 0.22 Ω. This both improves voltage-swing efficiency and reduces the extra distortion generated if the amplifier is erroneously biased into Class AB.[8] Thus even with low-impedance loads the $R_e$ voltage drop is very small, and insufficient to turn on a silicon power diode at realistic output levels.

Schottky diodes have a much lower forward voltage drop and might be useful here. Tests with 50 A diodes have been made but have so far not been encouraging in the distortion reduction achieved. A suitable Schottky diode costs at least as much as an output transistor, and two will be needed.

## The trouble with triples

In electronics, there is often a choice between applying brawn – in this case using multiple power devices – or brains to solve a given problem. The 'brains' option would be represented by a clever circuit configuration that gave the same results without replication of expensive power silicon.

The obvious place to start looking is the various output-triple topologies that have occasionally been used. Note that 'output-triples' here refers to

pre-driver, driver, and output device all in a local negative-feedback loop, rather than three identical output devices in parallel, which I would call 'tripled outputs'. Nomenclature is a problem.

In simulation, output-triple configurations do indeed reduce the gain-droop that causes large-signal nonlinearity. There are many different ways to configure output-triples. They vary in their general linearity and effectiveness at minimising large-signal nonlinearity.

The real difficulty with this approach is that three transistors in a local loop are very prone to parasitic and local oscillations. This is exacerbated by reducing the load impedances, presumably because the higher collector currents lead to increased device transconductance. This sort of problem can be very hard to deal with, and in some configurations appears almost insoluble. I have not studied this approach further.

## Loads below 4 Ω

So far I have concentrated on 4 Ω loads; loudspeaker impedances can sink lower than this, so I pursued the matter down to 3 Ω. One pair of *3281/1302* devices will deliver 50 W into 3 Ω for THD of 0.006% at 10 kHz, Figure 10. Two pairs of *3281/1302*s reduce this to 0.003% at 10 kHz, Figure 11. This is a very good result for such simple circuitry, and may be something of a record for 3 Ω linearity.

At this point it seems that whatever the device type, doubling the outputs halves the THD percentage for 4 Ω loading. The principle can be extended down to 2 Ω operation, but tripled devices are required for sustained operation at significant powers. Resistive losses are serious, so 2 Ω power output may be little greater than that into 4 Ω.

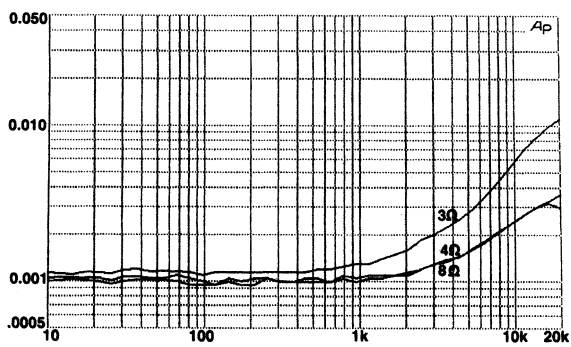



**Figure 10** *Distortion for* 3, 4 *and* 8 Ω *loads, single 3281/1302 devices,* 20 W/8 Ω, 40 W/4 Ω *and* 60 W/3 Ω.

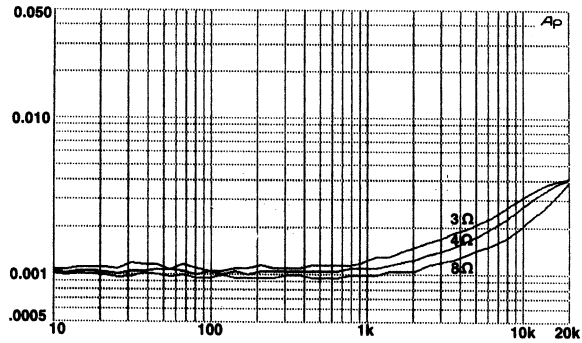



**Figure 11** *Distortion for* 3, 4 *and* 8 Ω *load, double 3281/1302 devices. Power as Figure 10.*

## Improved 8 Ω performance

It was wholly unexpected that the sustained-beta devices would also show lower crossover distortion at 8 Ω – but they do. What is more, the effect is again repeatable.

Possibly, whatever improves the beta characteristics has also somewhat altered the turn-on law so that crossover distortion is reduced; alternatively traces of large-signal nonlinearity, not visible in the THD residual, may have been eliminated.

Plot Figure 11 shows the improvement over the *MJ15024/25* pair; the 8 Ω THD at 10 kHz is reduced from 0.003% to 0.002%, and with correct bias adjustment, crossover artifacts are simply not visible on the 1 kHz THD residual.

The artifacts are only just visible in the 4 Ω case. To get a feel for the distortion being produced, and to set the bias optimally, it is necessary to test at 5 kHz into 4 Ω.

## Implementing the load-invariant concept

Figure 12 shows the circuit of a practical load-invariant amplifier intended for 8 Ω nominal loads with 4 Ω impedance dips. Its distortion performance is shown in Figures 6–11, depending on the output devices fitted.

Apart from load-invariance, this design also incorporates two new techniques from the thermal dynamics series.

The first technique greatly reduces time-lag in the thermal compensation. With a complementary-feedback pair output stage, the bias generator aims to shadow driver junction temperature rather than the outputs. A much faster response to power dissipation changes is obtained by mounting the bias generator transistor $Tr_8$ on top of driver $Tr_{14}$, rather than on the other side of the heat-sink. Driver heat-sink mass is thus largely decoupled from the thermal

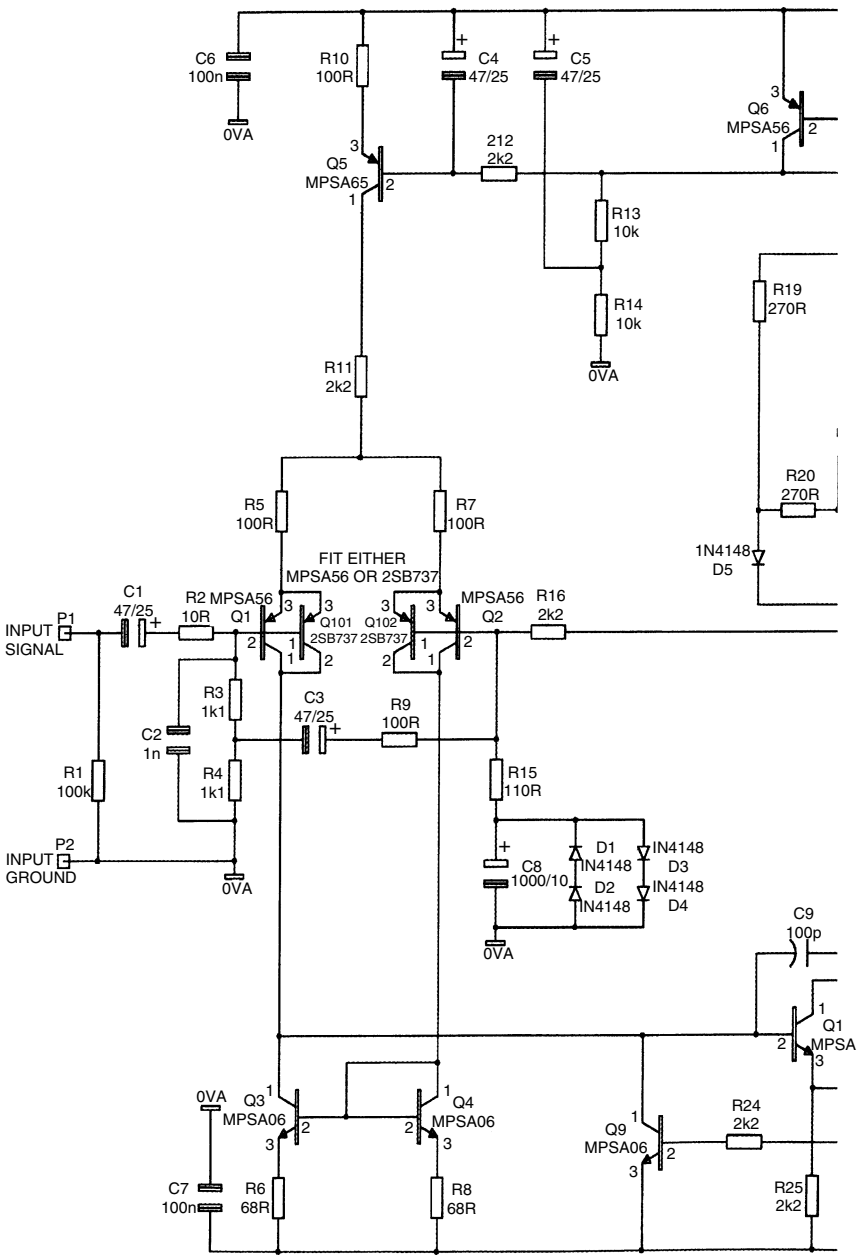



**Figure 12** *This load-invariant power amplifier is designed to keep performance constantly high as the loudspeaker impedance rises and falls with frequency. It is intended for* 8 Ω *nominal loads with* 4 Ω *impedance dips. Distortion, Figures 6–11, depends on output devices fitted.*

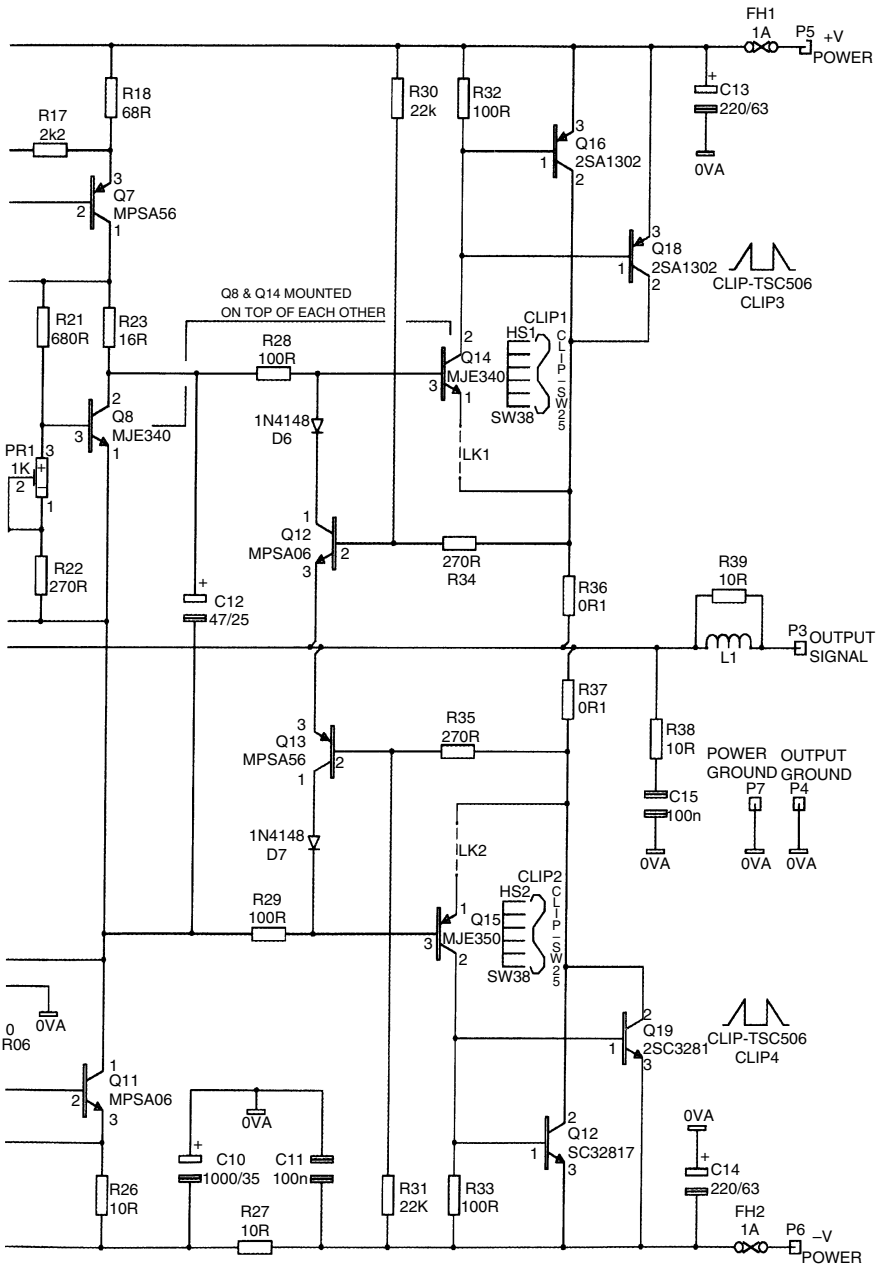



**Figure 12** *Continued*

compensation system, speeding up the response by at least two orders of magnitude.[9]

The second new technique is the use of a bias generator with an increased temperature coefficient, to reduce the static errors introduced by thermal losses between the driver and the sensor. Temperature coefficient is increased to $-4.0\,\text{mV}/^{\circ}\text{C}$.[10] Diode $D_5$ also compensates for the effect of ambient temperature changes.

The design is not described in detail because much of it closely follows the Blameless Class-B amplifier described Refs 1 and 11. Some features are derived from the Trimodal amplifier.[8] Most notable of these is the low-noise feedback network, with its requirement for input boot-strapping if a $10\,\text{k}\Omega$ input impedance is required. Single-slope *VI* limiting is incorporated for overload protection; see $\text{Tr}_{12,13}$.

As usual the global negative feedback factor is a modest 30 dB at 20 kHz.

## A point of departure

The improvements described here fit neatly into the philosophy of Blameless power amplifiers. The fundamental principle of the Blameless concept is that Distortion 3 should be the only significant distortion remaining. Distortions 1, 2 and 4–8 can all be reduced to negligible levels in straightforward ways.

For $8\,\Omega$ operation, the main nonlinearity left is crossover distortion, which seems to vary only very slightly with output transistor type.

As I hoped, the concept of a Blameless power amplifier is proving extremely useful as a defined point of departure for new amplifier techniques. Starting from the standard Blameless Class-B amplifier, I have derived:

- The pure Class-A power amplifier[12]
- The Trimodal A/AB/B amplifier[8]
- The load-invariant amplifier described here
- A further new design to be announced.

Note that Trimodal and new load-invariant amplifier are simple add-ons to the basic Blameless Class-B configuration. The Trimodal design adds a Class-A biasing subsystem, and the new amplifier grafts on extra – or improved – output devices.

## In summary

This study is incomplete in that the details of the large-signal nonlinearity mechanism remain incompletely understood, even though several practical methods for reducing it now exist. A detailed mathematical analysis would

probably get to the bottom of it, but a foot-long equation usually gives little physical insight.

My initial thoughts were that an amplifier could be considered as load-invariant if the rise in THD from 8 Ω to 4 Ω was less than some given ratio. For normal amplifiers the THD increase factor is from two to three times. The actual figure attained by the amplifier presented here is 1.2 times. I, for one, am prepared to classify this as 'load-invariant'. The ratio could probably be made even closer to unity by tripling the outputs.

Remember that this amplifier is designed for 8 Ω nominal loads, and their accompanying impedance dips; it is not intended for speakers that start out at 4 Ω nominal and plummet from there. Nonetheless, I hope it is some progress towards load-invariance, and that power amplifier design might have taken another small step forward.

## References


1. Self, D. '*Audio Power Amplifier Design Handbook*', Newnes, ISBN 0–7506–2788–3 (The Blameless amplifier concept, and more).
2. Self, D. 'Night thoughts on crossover distortion', *Electronics World*, November 1996, p 858.
3. Self, D. 'Distortion in power amplifiers: part 4', *Electronics World*, November 1993, p 929 (Distortion Number 3a: large-signal nonlinearity).
4. Self, D. 'Distortion in power amplifiers: part 6', *Electronics World*, January 1994, p 43, 44 (Distortion Number 4: VAS loading).
5. Self, D. 'Distortion off the rails', *Electronics World*, March 1995, p 201 (Supply rejection in power amplifiers).
6. Self, D. 'Distortion in power amplifiers: part 5', *Electronics & Wireless World*, December 1993, p 1010 (Large-signal nonlinearity and the Beta-droop effect. Feedforward diodes).
7. Self, D. 'FETs vs BJTs – the linearity competition', *Electronics & Wireless World*, May 1995, p 387 (The poor linearity of power FETs).
8. Self, D. 'Trimodal audio power: part I', *Electronics World*, June 1995, p 465 (Reduction of THD in Class AB by low $R_e$'s).
9. Self, D. 'Thermal dynamics in audio power: part II', *Electronics World*, June 1996, p 484 (Reduction of sensor delay by mounting on driver).
10. Self, D. 'Thermal dynamics in audio power: part III', *Electronics World*, October 1996, p 755 (Bias generator with increased temperature coefficient and ambient compensation).
11. Self, D. 'Distortion in power amplifiers: part 6', *Electronics & Wireless World*, January 1994, p 41 (The Blameless amplifier concept).
12. Self, D. 'Distortion in power amplifiers: part 8', *Electronics & Wireless World*, March 1994, p 229 (A Class-A power amplifier).

# 29 Common-emitter power amplifiers: a different perception?

*July 1994*

> Now and again articles appear in Electronics World that provoke thought. The sum-of-squares principle of Mr Williams sounded most intriguing, so I ran a few simulations to see if the linearity it could provide was an improvement over conventionality. This proved not to be the case, but it did lead to an interesting intellectual journey. As usually happens when I am evaluating an idea, enough work had been done for an chapter on the topic, and here it is.

When I read Michael William's intriguing article *Making a Linear Difference to Square-Law fets*,[1] I was attracted by the prospect of applying it to an audio power output stage. I found the phrase 'curvilinear class A' particularly appealing.

The basic concept of the difference-of-squares is not new, as several correspondents to $EW + WW$ have pointed out.[2,3] Another early reference (1949) to the quarter-squares principle can be found in the monumental MIT Radiation Lab series on radar techniques.

Mr William's basic circuit is shown in Figure 1, and the first problem to overcome in applying it for audio power is that the wanted output is the difference of two currents whereas hard-bitten amplifier designers are more used to a low impedance voltage output. Note that with the usual enhancement-mode power fets, if $V_1$, $V_2$ are a.c. sources only, and carry no d.c. bias, then $V_b$ will have to establish point $M$ some volts below ground. No doubt something could be done with industrial-sized current-mirrors, but it struck me that the circuit could be rearranged as Figure 2, by making use of complementary devices. We now need two bias voltages $V_{b1}$, $V_{b2}$, and the positioning of the two signal sources $V_1$, $V_2$ on opposite rails looks

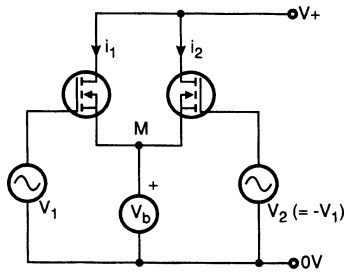



**Figure 1** *The original Williams circuit; the output required is the difference between* $i_1$ *and* $i_2$.

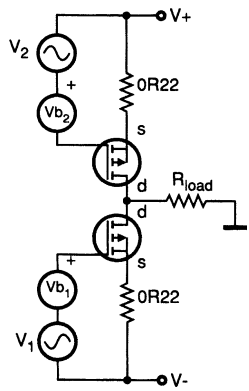



**Figure 2** *The* $i_1$, $i_2$ *subtraction carried out by inverting the polarity of one of the FETs. Two bias voltage generators are now needed.*

a little awkward, but at least the current-difference will be mathematically perfect, if Kirchhoff has anything to say on the matter.

So far so good. We now have a single current output $i_{out}$. But is this any use for driving loudspeakers? I am assuming that current-drive of speakers is not the final goal; I appreciate that this can be made to work, and promises some tempting advantages in terms of reducing bass-unit distortion.[4] My immediate reaction to Figure 2 was no, it can't work, because with a high impedance output, the output stage gain will vary wildly with load impedance making the amount of NFB applied a highly variable quantity. It would also appear that any capacitive loading of this high-impedance node would generate an immediate output pole that would make stable compensation a waking nightmare.

However, just as I was discarding the notion, it occurred to me that the structure in Figure 2 looks very much like the bipolar common emitter (CE) stage in Figure 3. This is widely used in low voltage opamps because the low saturation voltage allows a close approach to the rails.[5] The more usual emitter follower type of opamp output is usually called a CC or

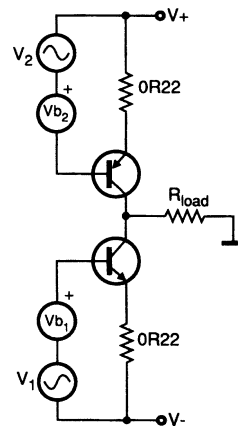



**Figure 3** *The bipolar version of Figure 2, as used in many low-voltage opamps and Walkman output amplifiers.*

common-collector stage. It is highly probable that the widest application of these voltage-efficient CE configurations is in the headphone amplifiers of personal stereos.

At about the same time I encountered a paper by Cherry[6] which pointed out that, so long as NFB is applied, the output impedance of such a stage can be as low as for the usual voltage follower type output. Cherry's paper is dauntingly mathematical, so I will summarise it thus. The vital point about using NFB to reduce the output impedance of an amplifier is that the amount of NFB applied must be calculated *assuming that the open-loop case is unloaded*. This condition looks unfamiliar, because the average amplifier usually has a fairly low output resistance even when open-loop, due to its output follower configuration, and so the loaded/unloaded distinction makes only a negligible difference when calculating the reduction of output resistance by NFB.

Using this condition, Cherry shows that output impedance of a CE stage should be exactly equivalent to the usual CC stage, when the global NFB is applied. I appreciate that this result is counter-intuitive; it looks as though the current output version must have a higher output impedance, even with NFB, but it appears not to be so. Doubters who are unafraid of matrix algebra should consult Cherry's paper.

## Topology to the test

Nonetheless, before reaching for the power fets, I felt the need for further reassurance that a CE output stage was workable. There are several low voltage opamps that use the CE output topology, so it seemed instructive

to provoke one of these with some output capacitance and see what happens. A suitable candidate is the Analog Devices *AD820*, which has a BJT output stage looking like Figure 3 and provides all you need for CE experimentation in one 8-pin package.[7]

My practical findings were that the opamp works well, and while THD may not be up to the very best standards, it was happy with varying load resistances, proved stable with capacitors hung directly on the output, and was relaxed about rail decoupling. Once again, so far, so good.

By this stage, the quarter-squares principle was slipping somewhat into the background. My attention was focusing on the possibilities of a BJT power output stage something like Figure 4, which shows the addition of drivers and emitter resistors to make the circuit more practical. A good output swing is facilitated by the inward-facing driver arrangement. In a conventional emitter follower output the need to leave the drivers room to work in further reduces output swing.

Figure 4 could be configured into something like a normal Class-B amp, except that the novel use of a CE output stage would allow greater efficiency than usual because there would be the low $V_{ce(sat)}$ drops mentioned above. Also the crossover behaviour would presumably be different from a normal CC output, and quite possibly better, or at least more easily manipulated.

In a previous article[8] I tried to demonstrate that for an amplifier in which all the easily manipulated distortion mechanisms had been suitably dealt with, the low frequency THD was below the noise when driving an 8 Ω load. This without large global feedback factors: 30 dB at 20 kHz is quite adequate.

At high frequencies (say above 2 kHz) the distortion is easily measurable, and almost all of its results from crossover effects in the output

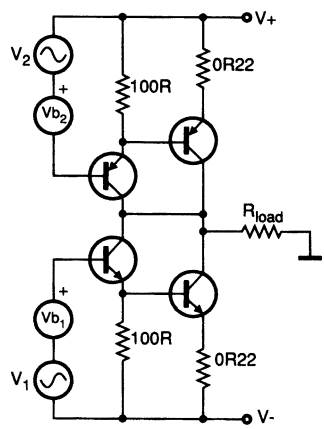



**Figure 4** *A practical circuit based on Figure 3. Drivers and emitter-resistors have been added.*

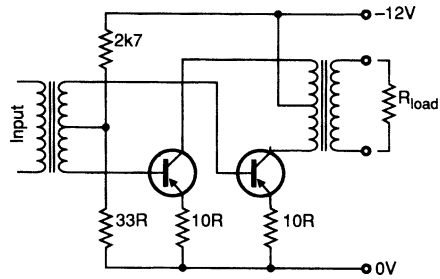



**Figure 5** *A rather old-fashioned CE amplifier: the transformers are expensive but avoid the need for complementary devices.*

stage. Since NFB typically falls with frequency, these high-order harmonics receive much less linearisation. This is why any technique that promises a reduction in basic crossover nonlinearity is of immediate interest to those concerned with power amplifier design.

I began to think that Mr Williams had opened up a whole new field of audio amplification; each conventional CC output stage would have its dual in CE topology, perhaps with new and exciting characteristics.

The next stage of the investigation was more sobering. There was a familiarity about CE output stages. Readers old enough to recall paying 30 shillings for their first OC72 will recognise Figure 5 as the configuration used almost universally for low power audio output for many years when there was no such thing as a complementary device. Transformers provide one way to make a push-pull output. At first sight bias voltage $V_b$ looks as if it will be far too low but bear in mind these are germanium transistors. Note the upside-down format of the circuit which is typical of the period. The circuit values are appropriate for an output of about 500 mW.

While it is perhaps not obvious, this is the equivalent of Figure 3. The need for an npn is avoided by using phase inversions in the transformers. So clearly CE output stages were not as rare and specialised as I thought; however they might still have handy distortion properties that were not obvious in the long-gone days of transformer coupling.

## Adding Spice to the investigation

The next step was Spice simulation of the practical BJT output circuit in Figure 4: Figure 6 shows how the device currents vary in a relationship that looks ominously like classic Class-B. Somehow I was expecting more overlap of conduction. The linearity results are presented in Figure 7 as a plot of incremental gain versus output voltage for varying loads, as in the *Distortion In Power Amplifiers* series.[8]

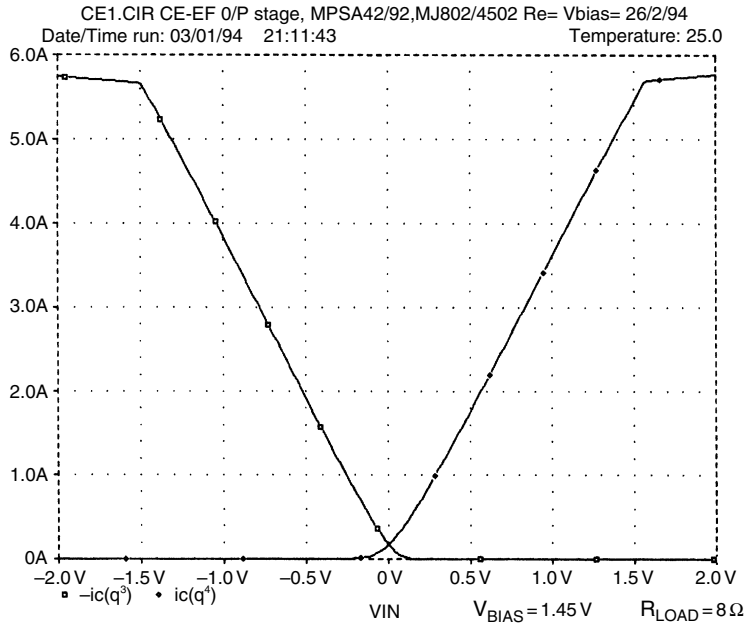



**Figure 6** *BJT Collector currents in Figure 4 driving an* $8\,\Omega$ *load.*

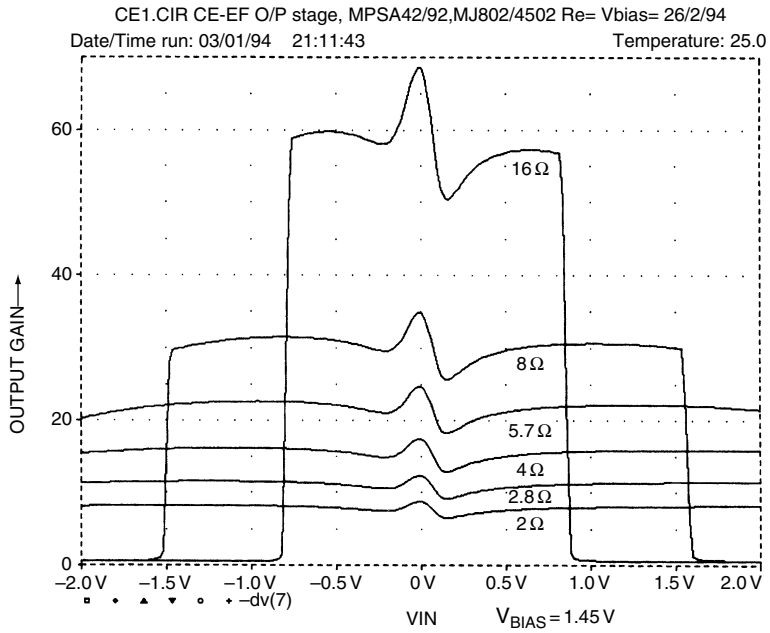



**Figure 7** *Gain linearity of Figure 4, various load resistances (BJT).*

The first obvious difference is that stage gain, instead of staying close to unity, varies hugely with load impedance – pretty much what we expect from a CE stage operating open-loop. Note that the X-axis is $V_1$ ($V_2 = -V_1$ to induce push-pull operation) and so represents the input voltage only rather than both input and output as before. Multiplying this input voltage by the gain taken from the Y-axis gives the peak output voltage swing. The vertical gain drop-offs that indicate clipping move inwards with higher load impedances because of the greater output gain rather than through any hidden limitation on output swing.

Figure 8 shows the effect of varying the bias, and hence quiescent current, for an 8-Ω load.

This circuit certainly works, but somehow the linearity results seem depressingly familiar. There is the same gain-wobble at crossover we have seen *ad nauseam* with CC output stages, and once again there is no bias setting that removes or significantly smooths it out. As before, the usual falling-with-frequency NFB will not deal with this sort of high-order distortion very effectively, leading to a rise in THD above the noise in the upper audio band.

In fact, the characteristics look so suspiciously similar to the standard emitter-follower CC stage, that it began to belatedly dawn on me they might actually be the same thing . . .

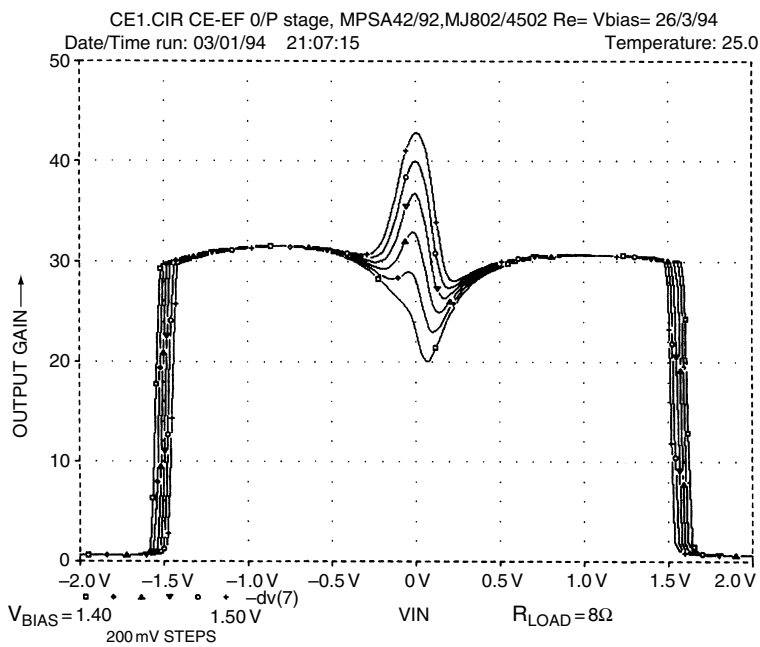



**Figure 8** *Gain linearity of Figure 4 for various bias voltages, load is* 8 Ω *(BJT).*

Figure 9 shows the final stages of this conceptual hejira. Figure 9(a) shows the simplified circuit of Figure 3 with the power supplies V+, V- included; they no doubt come from a mains transformer so we can float them at will, and it seems quite in order to pluck them from their present position and put them in the collectors of the output devices instead. All the other supplies shown are equally without ties forming an independent unit with the associated transistor and emitter resistor $R_e$. Thus they cannot affect device currents. Since there is only one ground reference in the circuit, it is also a legitimate gambit to put it wherever we like, which in this case is now the opposite end of the load $R_l$. (See Ref. 9 for another example of this manoeuvre). This gives us the unlikely looking but functionally equivalent circuit in Figure 9(b).

A purely cosmetic rearrangement of Figure 9(b) produces Figure 9(c), which is topologically identical, and reveals that the new output stage is . . . a CC stage after all. Figure 9(d) shows the standard output.

The only true difference between the 'CE' stage and the traditional CC stage is the arrangement of the two bias voltages $V_{b1}$, $V_{b2}$. In a conventional CC stage, the output bases or gates are held apart by a single fixed voltage, shown here as $V_{b1}$ and $V_{b2}$ connected together. This rigid 'unit' can be regarded as driven with respect to the output rail by the signal source $V_{sig}$, representing the difference between input and output of the stage. Normally, of course, it is more useful to regard the earlier circuitry as generating a signal voltage with respect to ground.

In contrast to Figure 9(d), Figure 9(c) has two bias voltage generators, and the consequence of this is that voltage drops in the emitter resistors $R_e$ are not coupled across to the opposite device by the bias voltage. This does not seem to offer immediately any magical stratagems for reducing the gain deviation around crossover, and creates the need for two drive voltages referenced about the output rail. This should be fairly easy to contrive, but is bound to be more complex than the traditional method.

## Squaring the circle

Having gone through these manipulations, it is time to reconsider fets and the quarter-squares approach, knowing now that we are dealing with something very close to a standard power-amp configuration. To underline the point. Figure 10 shows the gain characteristics for the circuit of Figure 2, using *2SK135/2SJ50* power fets. Note the very close resemblance to a conventional source follower.[8]

As Mr Williams points out, the $V_{gs}/I_d$ characteristic curve for power fets may follow a square law at low currents, but it is more or less linear at high ones, and this appears to rule out any simple approach to 'curvilinear class A'. For the fets I used, the 'square lawish' region is actually tiny,

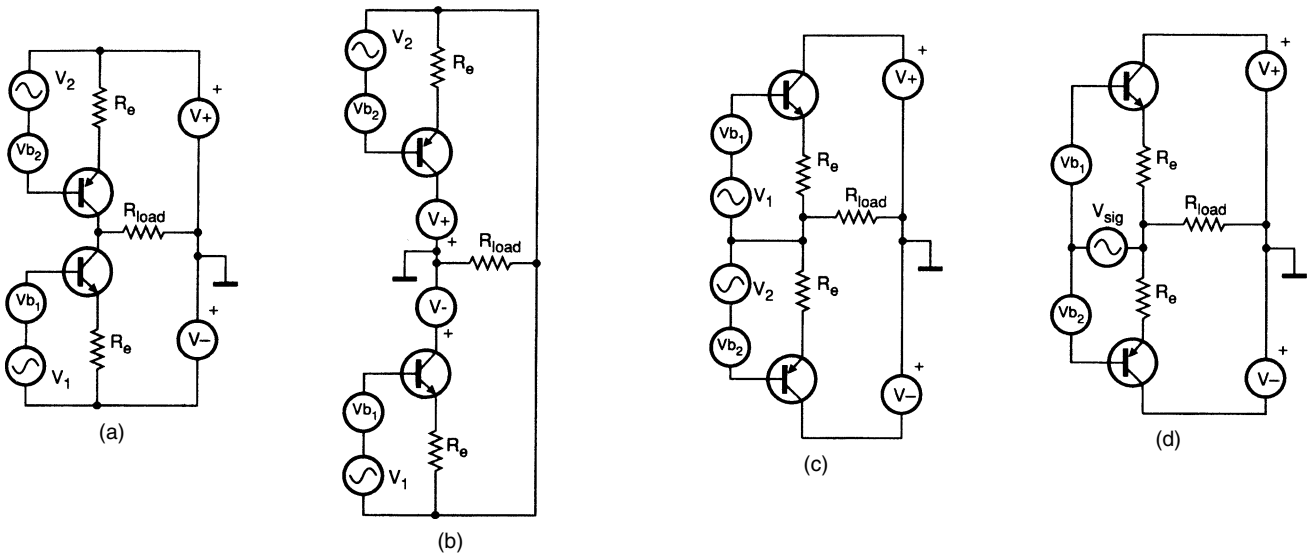



**Figure 9** *Showing how our experimental CE amplifier turns out to be a more or less conventional CC amp when turned inside out.*

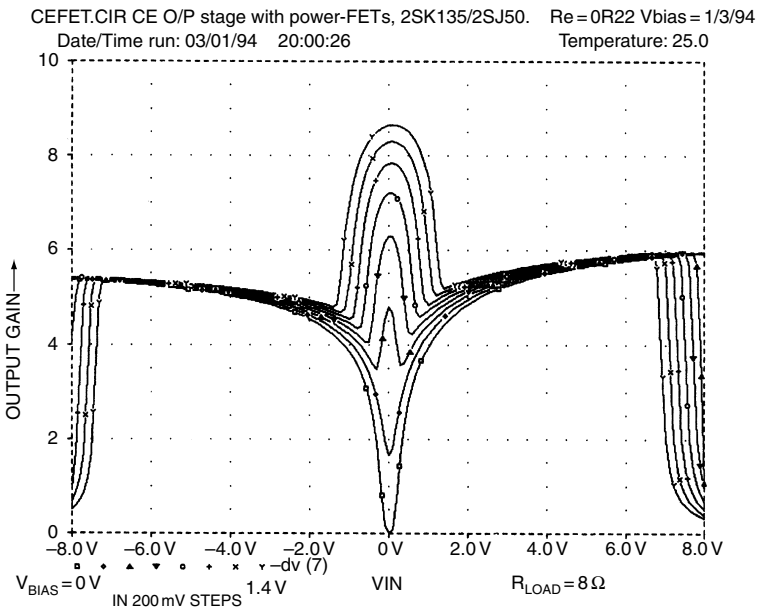



**Figure 10** *The gain linearity of the FET circuit in Figure 2 for various bias voltages. This looks very similar to a conventional source-follower output stage.*

being roughly between 0 to 80 mA which is of limited use for a power stage. In so far as second-harmonic cancellation occurs at all, it is in the crossover region where, without this effect, the central gain deviations would probably be greater than they are.

As I can see, the quarter-squares concept is already in use in most FET power amplifiers in heavy disguise but only operational in the crossover region. If this idea is to be pursued further, we need a true square-law output device. Since there is no such thing, it would need to be realised by some kind of law-synthesis circuitry. If amplifier distortion needs reducing below the tiny levels possible with relatively conventional techniques, there are probably better avenues to explore.

## References

1. Williams, M. 'Making a linear difference to square-law fets', *EW + WW*, January 94, p 82.
2. Brown, S F. Letters, *EW + WW*, March 94, p 247.
3. Owen, D, Letters, *EW + WW*, March 94.
4. Mills and Hawksford, 'Transconductance power amplifier systems, current-driven loudspeakers', *JAES* Vol 37 (10), October 1989.

5. Fonderie and Huijsing, *Design of Low-Voltage Bipolar Operational Amplifiers*, Kluwer Academic Publishers, 1993, Chapter 3.
6. Cherry and Cambrell, '*Output Resistance & Intermodulation Distortion of Feedback Amplifiers', JAES*, Vol 30 (4), April 82.
7. Jung and Wong, 'High-performance ICs in single-supply analog circuits', *Analog Dialogue* #27–2 (1993) p 16.
8. Self, D. 'Distortion in power amplifiers: Part 4'. *EW + WW*, November 93, p 932.
9. Baxandall, P. 'Symmetry in Class B', Letter, *Wireless World*, September 1969, p 416.

# 30 Few compliments for non-complements

*September 1995*

A further article by another author, Mr Olsson, also caused me to begin my own investigations, not least because few performance details were given in the original. The concept might have worked, but it did not, so all I could really do was say so, adding some more positive material on two-stage amplifiers in general. They turned out to be rather awkward things.

I want to stress that the last thing intended in these 'reactive' articles is discourtesy to the original authors. However, what I do is Science, which is based on Truth, and the latter has proverbially no respect for persons.

Bengt Olsson's most interesting article on quasi-complementary FET output stages[1] prompted me to examine how his proposed configuration works. Investigations showed that his scheme changes not just the output stage but the entire structure of the amplifier, and it presents some intriguing new design problems.

## An alternative architecture

Nearly all audio amplifiers use the conventional architecture I have analysed previously.[2] There are three stages, the first being a transconductance stage, i.e. differential voltage in/current out, the second is a transimpedance stage i.e. current in/voltage out and lastly a unity-gain output stage, Figure 1(a).

Clearly, the second stage has to provide all the voltage gain and is therefore formally named the voltage amplifier stage (VAS). This architecture has several advantages. A main benefit is that it is straightforward to arrange things so that the interaction between stages is negligible. For example, there is very little signal voltage at the input to the second stage, due to its current-input

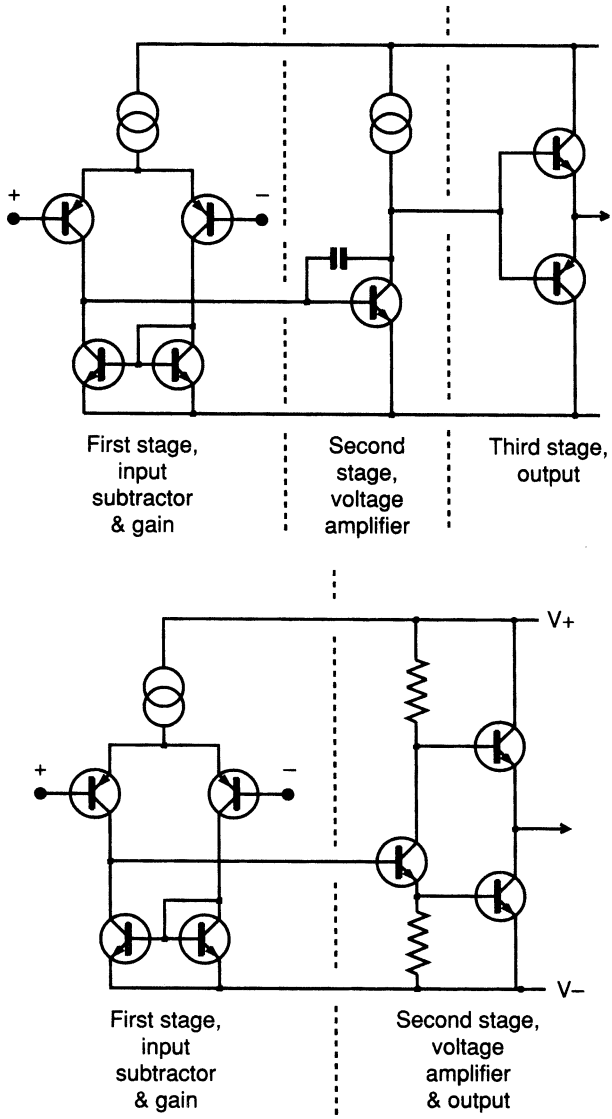



**Figure 1** *Conventional three-stage amplifier architecture, and two-stage architecture advocated by Olsson.*

nature. This results in very little voltage on the first stage output, which in turn minimises phase-shift and possible Early effect.

In contrast, the architecture presented by Olsson is a two stage amplifier, Figure 1(b). The first stage is once more a transconductance stage, though now without a guaranteed low impedance to accept its output current. The

second combines voltage amplifier stage and output stage in one block. It is inherent in this scheme that the voltage amplifier must double as a phase-splitter. This results in two dissimilar signal paths to the output, and it is not at all clear that trying to break this block down further will assist a linearity analysis. The use of a phase-splitting stage harks back to valve amplifier days, when it was essential due to the lack of complementary valve technology.

Since the amount of linearising global feedback available depends upon amplifier open-loop gain, the way in which the stages contribute to this is are of great interest. The normal three-stage architecture always has a unity-gain output stage – unless you really want to make life difficult for yourself. As a result the total forward gain is simply the product of the transconductance of the input stage and the transimpedance of the voltage amplifier stage. Transimpedance is determined solely by the Miller capacitor $C_{dom}$, except at very low frequencies.[3]

Typically, the feedback factor at 20 kHz will be 25–40 dB. It will increase at 6 dB per octave with falling frequency until it reaches the dominant pole frequency $P_1$, when it flattens out. What matters for the control of distortion is the amount of negative feedback, NFB, available, rather than the open-loop bandwidth, to which it has no direct relationship.

In my *EW + WW* Class-B design, input stage $g_m$ is about 9 mA/V, and $C_{dom}$ is 100 pF, giving a feedback factor of 31 dB at 20 kHz. In other designs I have used as little as 26 dB at the same frequency with good results.

Arranging the compensation of a three-stage amplifier can be relatively simple. Since the pole at the voltage-amplifier stage is already dominant, it can be easily increased to lower the h.f. negative-feedback factor to whatever level is considered safe. The local negative feedback working on the voltage amplifier has an invaluable linearising effect. I am aware that some consider there are better ways to perform this sort of compensation, but the Miller approach is so far the most stable method in my experience.

## Fewer stages, more complexity?

Paradoxically, a two-stage amplifier may be more complex in its gain structure than a three-stage. Forward gain depends on the input-stage $g_m$, the input-stage collector load, and the gain of the output stage, which will be seen to vary in a most unsettling manner with bias and loading. Input-stage collector loading plays a part here since the input stage cannot be assumed to be feeding a virtual earth.

Choosing the compensation is also more complex for a two-stage amplifier. The voltage-amplifier/phase-splitter has a significant signal voltage on its input. Usually, the pole-splitting mechanism enhances Nyquist stability by increasing the pole frequency associated with the input stage collector.

But because of the relatively high voltage on the voltage-amplifier/phase-splitter, the pole-splitting mechanism is no longer effective.

This may be why Olsson's circuit uses a cascoded input stage comprising $Tr_{6,7}$ in his original circuit, Figure 11. This presents the input device collectors with a low impedance and prevents a significant collector pole. Another valid reason is that it also allows the use of high-beta low-$V_{ce}$ input transistors, which minimise output d.c. offset due to base current mismatch. This is usually much larger than the d.c. offset due to $V_{be}$ mismatch.

Such an input cascode can also improve power-supply rejection as it prevents Early effect from modulating the subtractive action of the input pair.

Simple calculation gives the $g_m$ of Olsson's amplifier as 16 mA/V, but the effective gain of the next stage seems much more difficult to equate. A full *PSpice* simulation of the complete amplifier with an 8 Ω load shows that the feedback factor is 36 dB up to 300 Hz. It then rolls off at the usual 6 dB/octave, until it passes through OdB at about 20 kHz. This 36 dB represents much less feedback than the three-stage version. It indicates that $C_{dom}$, notionally connected between drain and gate of $M_3$, must be comparatively very large at approximately 3 nF.

Specified internal capacitances of $M_3$ are certainly orders of magnitude larger than those of an equivalent bipolar device – they vary from sample to sample and also with operating conditions such as $V_{ds}$. These unwanted variations would appear to make stable and reliable compensation a difficult business.

The low-frequency feedback factor is about 6 dB less with a 4 Ω load, due to lower gain in the output stage. However, this variation is much reduced above the dominant pole frequency, as there is then increasing local negative feedback acting in the output stage.

## Devices and desires

In his opening paragraph, Olsson says that the symmetry of complementary transistor output stages is theoretical rather than practical. Presumably he is referring only to power-fets, as suitable pairs of bipolar devices, such as Motorola *MJ802/MJ4502* – old favourites of mine – exhibit excellent symmetry.[4] Admittedly, the two devices are not exact mirror-images, but the asymmetries are small enough for even-order harmonic generation in the output stage to be negligible. This is surely what counts.

This symmetry does not hold for power-fets however, and so it may be that some of Olsson's concern with symmetry flows from an initial decision to use fets. I find it difficult to understand why power fets in particular suffer from so many mis-statements. It is still confidently held that fets are more linear than bipolars, although the opposite is certainly the case when the two types of device are used in normal Class-B output stages.

Similarly, FET robustness is often exaggerated, the devices being prone to summary explosion under serious parasitic oscillation. Mercifully bipolars are not. In particular, I find it hard to understand Olsson's contention that FET parameters are predictable – they are notorious for being anything but.

From one manufacturer's data, namely Harris, the $V_{gs}$ for the *IRF240* varies between 2 and 4 V for an $I_d$ of 250 μA – a range of two to one. In contrast the $V_{be}/I_c$ relation in bipolars is fixed by a mathematical equation for a given transistor type. The exponential relationship may be regarded as more non-linear than the partially-square-law FET $V_{gs}/I_d$ relationship but it is dependable, and gives a much higher transconductance. This can always be traded for linearity by introducing local negative feedback.

## Output considerations

Figure 2 shows the basic output configuration I have investigated – I have not examined the 'anti-saturation' schemes intended to provide the output fets with extra gate drive.

My first discovery is that the voltage-amplifier stage/phase-splitter does not have to be a FET. Replacing it with a bipolar junction transistor, for example *MPSA92*, gives almost identical results. I have used $M_{1,2}$ etc. rather than $Tr_{1,2}$ for FET designations as this preserves consistency with the *PSpice* output.

This output stage configuration is totally different in operation from the conventional Class-B stages discussed in Ref. 1. It is a hybrid

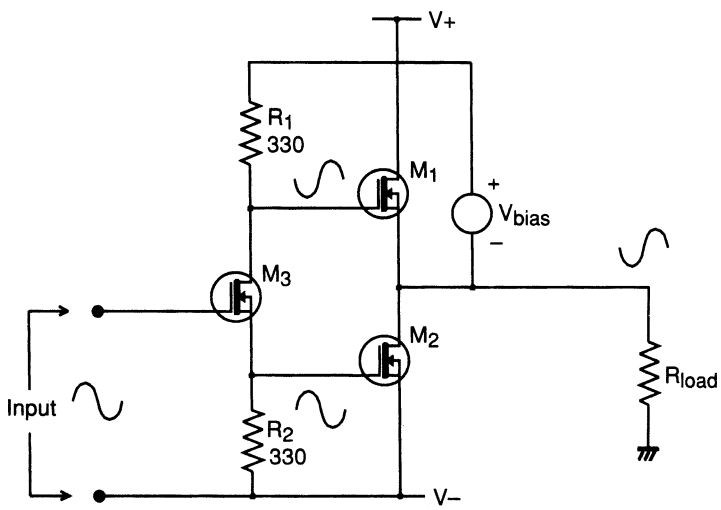



**Figure 2** *Basic FET output stage used for simulations of Figures 3, 4 and 5. Transistors* $M_{1,2}$ *are IRF240 and* $M_3$ *is IRF710.*

common-drain/common-source configuration, or, in bipolar speak, a common-collector/common-emitter (cc–ce) stage. In this sort of output, the upper emitter-follower has a common-emitter active-load. This load may or may not deliver an appropriate current into the node it shares with the upper device.

The input-voltage/output-current relationship for the upper and lower devices will be different, as a result of the two dissimilar paths to the output. This means that while such a stage can always be biased into Class-A by increasing the quiescent enough[5] there is every likelihood that it will be an inefficient kind of Class-A. It will deviate seriously from the constant-sum-of-currents condition that distinguishes classical Class-A.[6]

Lower quiescents tend to give a depressingly non-linear and asymmetrical Class-AB. In general there is no equivalent at all to standard Class-B, where the symmetry of the configuration – rather than anything else – allows both output devices to be biased to the edge of conduction simultaneously.

In general, cc–ce stages generate large amounts of even-order distortion, due to their inherent asymmetry. I appreciate however that avoiding this is one of the prime purposes of Olsson's circuit.

In Figure 2, the position of $V_{bias}$ causes a form of bootstrapping, and I can confirm that driving it from the output is essential to make the scheme work.

Figure 3 shows drain current in each output FET when driving an 8 Ω load, with $V_{bias}$ stepped, as simulated by *PSpice*. Compared with conventional Class-B, the drain currents cross over smoothly. This seems intuitively a good idea, and has been recommended by several writers, but in fact a smooth-looking current crossover does not guarantee a linear composite gain characteristic.

Figure 4 shows the input/output characteristic for the stage. You can see that the lower $V_{bias}$ levels produce a sigmoidal transfer function, with gain falling off in the crossover region. This gross output distortion is much greater than that given by a normal Class-AB stage. It suggests that it is only practical to run the stage in full Class-A, i.e., the straight line at the right of the plot. I do not recall a mention of this point in the original article.

Quiescent current needed to achieve this is about 2 A. The desirability of Class-A operation is reinforced by the incremental gain plot in Figure 5. It is clear that the gain variations are serious for lower $V_{bias}$, and do not augur well for the closed-loop distortion performance. Only the rightmost Class-A gain characteristic has a clear flat portion over its operational range.

It is true that the drain currents in this stage are symmetrical – but the quiescent required to remove the sharp-eared 'Batman' effect in Figure 5 is so high that the amplifier is working entirely in Class-A. The symmetry of the circuit means only that when distortion is produced, it will be predominantly odd-order, which is not normally considered a good thing from the subjective point of view. To keep the stage linear into 4 Ω loads

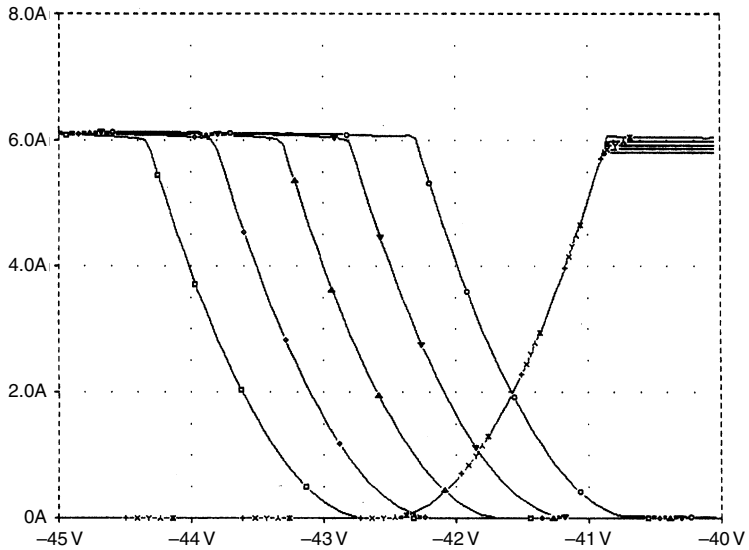



**Figure 3** *Drain current in each output FET when driving an* 8 Ω *load, with* $V_{bias}$ *stepped. As* $V_{bias}$ *increases the* $I_d(M_1)$ *line moves to the right and overlaps more with the* $I_d(M_2)$ *line. In each case symmetry exists about the intersection of the two* $I_d$ *lines. Values for* $V_{bias}$ *are* 7.5, 8, 8.5, 9 *and* 9.5 V.

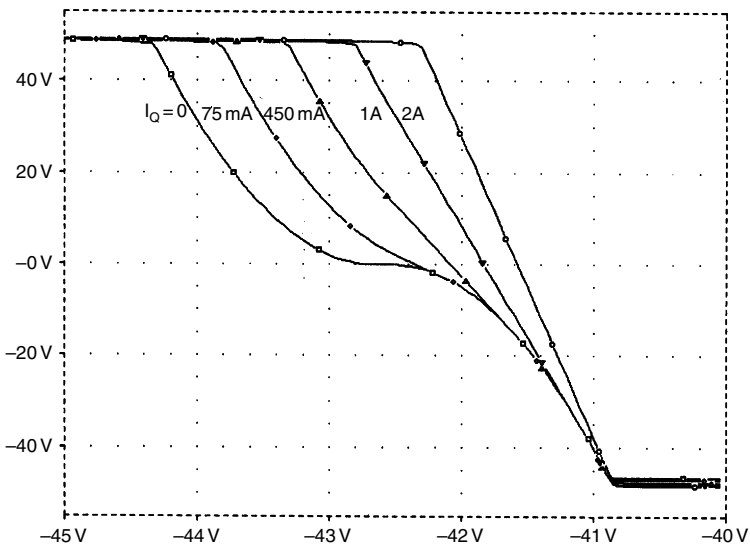



**Figure 4** *Output transfer functions for stepped bias voltage. For lower bias, the characteristic is sigmoidal – S-shaped, unlike a conventional Class-B stage. Values for* $V_{bias}$ *are* 7.5, 8, 8.5, 9 *and* 9.5 V.

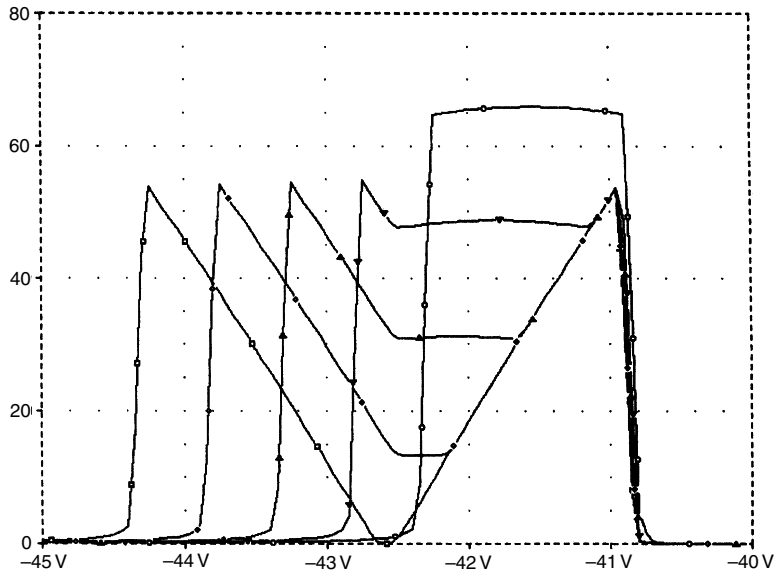



**Figure 5** *Incremental gain plot for the Olsson stage, with* $V_{bias}$ *stepped. The* –AB *mode (curve B) shows serious gain variations and therefore poor linearity.*

would demand a quiescent of 2.9A. It is interesting that this is not twice the current for the 8 Ω case.

There are other significant differences from the usual voltage-follower configuration, which if nothing else has a stage gain reliably close to unity. Olsson's output stage gain varies with $V_{\text{bias}}$ adjustment – even when in Class-A – and also varies strongly with load impedance. This would seem to make reliable compensation a difficult business, but in the complete amplifier this variation is probably only significant below the dominant-pole frequency $P_1$.

Figure 6 is a comparison between the Olsson configuration and three conventional stages, all biased to drive an 8 Ω load in Class-A. Traces 1 and 2, at the top, are bipolar-emitter follower and complementary-feedback-pair stages. These produce the usual linearity and close approach to unity gain.

The curved lower trace, 4, is a conventional complementary FET output. Trace 3 is the Olsson stage, with its gain of about 65 times normalised to fit in between the other curves. It shows stronger curvature than the bipolar stages, and despite everything, is actually less symmetrical than the usual output stages. I think this is inherent in the circuit's lack of symmetry about the output line.

I hope this article is a fair analysis of the proposed configuration, and that I have not made any serious misinterpretations of Mr Olsson's intentions.

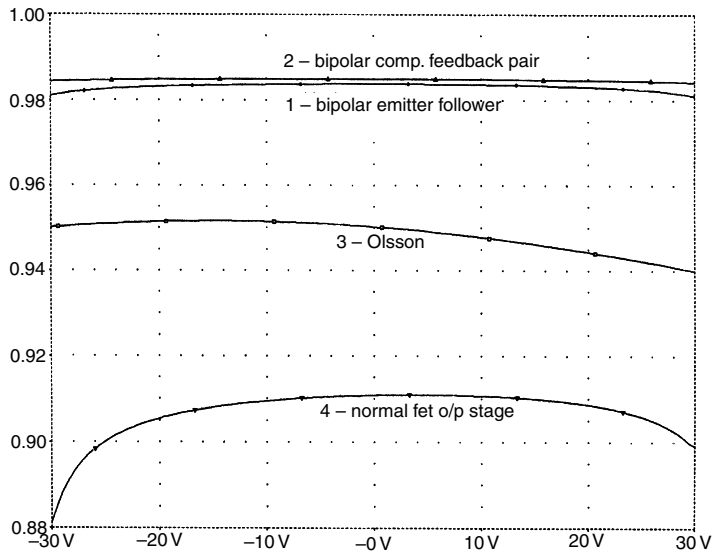



**Figure 6** *Curve 3 shows incremental gain for the Olsson stage, driving* 8 Ω *in Class-A. Equivalent plots for conventional Class-A bipolar emitter follower (curve 1), bipolar complementary feedback pair (curve 2) and a normal complementary FET stage (curve 4) are shown for comparison.*

I also trust that it will not be taken as purely destructive criticism, for that is not my intention. I have to conclude that the configuration appears to require a very high quiescent current for linear operation, and has only a limited amount of negative feedback available to correct output stage distortions. Any deeper investigation would need to be encouraged by some promise that the Olsson configuration can deliver substantial benefits, and as far as my analysis goes, this does not seem likely.

## References

1. Olsson, B. 'Better audio from non-complements?' *EW + WW*, December 1994, pp 988–992.
2. Self, D. 'Distortion in power amplifiers: Part 1', *EW + WW*, August 1993.
3. Self, D. 'Distortion in power amplifiers: Part 3', *EW + WW*, October 1993.
4. Self, D. 'Distortion in power amplifiers: Part 4', *EW + WW*, November 1993.
5. Hood, J. L. 'Simple Class-A amplifier, *Wireless World*, April 1969.
6. Self, D. 'Distortion in power amplifiers: Part 7, *EW + WW*, March 1994.

# 31 Loudspeaker undercurrents

*February 1998*

I must admit that when I first read that exotic waveforms that increased the current demands of speakers by several times could be thought up, I was a bit sceptical. Well, actually, I was very sceptical. However, I was wrong to be so untrusting. Simulating a very simple analog of a loudspeaker showed that transient currents of considerable size can be made to flow if careful adjustments are made to the stimulus waveform.

However, the implications of this for practical amplifier design are much less clear. The square-edged waveforms required to induce current-enhancement do not resemble speech or music at all, and in the intervening years since the article was published it has gradually become clearer that they do not exist in real life speech and music. It therefore seems that the extent to which such alarming current-enhancement occurs is negligible, and it need not be considered in the design process. Timely support for this view comes from an article called 'Current Affairs' by Keith Howard, in Hi-Fi News (Feb 2006) which finds no evidence of current-enhancement at all.

It is easy to show that the voltage/current phase shifts in reactive loudspeaker loads increase the peak power dissipation in a cycle, using sine wave test signals of varying frequency.[1] The effect of this on device selection in output stages is complicated by the inability to treat power ratings as average power, for as far as safe-operating areas are concerned, low audio frequencies count as d.c.

But sinewave studies do not give insight into what can happen with arbitrary waveforms. When discussing amplifier current capability and loudspeaker loading, it is often said that it is possible to synthesise special waveforms that provoke a loudspeaker into drawing a greater current than would at first appear to be possible. This is usually stated without further

explanation. Since I too have become guilty of this,[1] it seemed time to make a quick investigation into just how such waveforms are constructed.

The possibility of unexpectedly big currents was raised by Otala,[2] and expanded on later.[3] But these information sources are not available to everybody. The effect was briefly demonstrated in *EW* by Cordell,[4] but this was a long time ago.

## Speaker model

Figure 1 is the familiar electrical analogue of a single speaker unit. Component $R_c$ is the resistance and $L_c$ the inductance of the voice coil. In series, $L_r$ and $C_r$ represent the resonance of cone mass and suspension compliance, while $R_r$ controls the damping. These three components model the impedance characteristics of the real electromechanical resonance.

Voice-coil inductance is 0.29 mH, and coil resistance 6.8 Ω. These figures are typical for a 10 in bass unit of 8 Ω nominal impedance. Measurements on this load will never show an impedance below 6.8 Ω at any frequency. This makes it easy to assume that the current demands can never exceed those of a 6.8 Ω resistance. This is not true.

To get unexpectedly high currents moving, the secret is to make use of the energy storage in the circuit reactances, by applying an asymmetrical waveform with transitions carefully matched to the speaker resonance.

## Simulating the effects

Figure 2 shows PSpice simulation of the currents drawn by the circuit of Figure 1. The rectangular waveform is the current in a reference 8 Ω

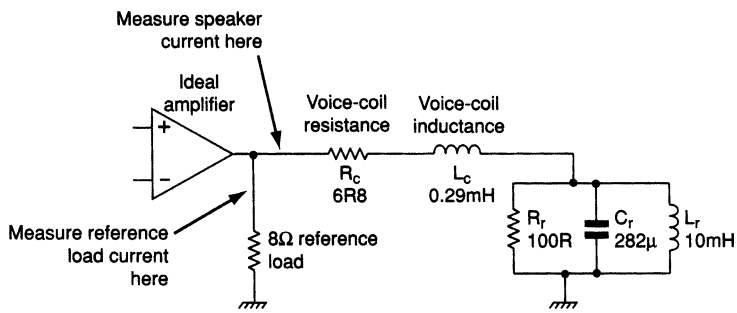



**Figure 1** *Equivalent circuit of single-unit speaker. Although the nominal impedance is* 8 Ω*, the coil resistance* $R_c$ *is only* 6.8 Ω*. This is based on a successful commercial 10 in bass unit. The LCR network on the right simulates the cone mass/suspension-compliance resonance.*

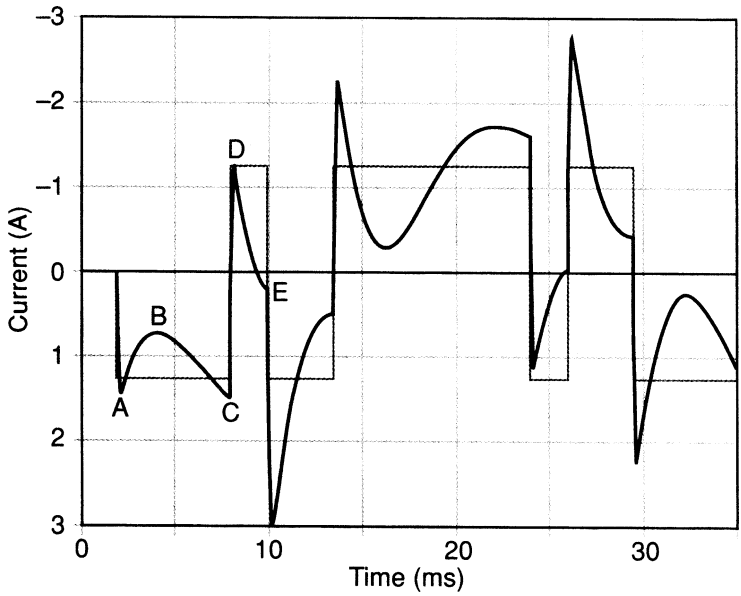



**Figure 2** *Asymmetrical waveforms to generate enhanced speaker currents. The sequence ABCDE generates a negative current spike; to the right, the inverse sequence produces a positive spike. The rectangular waveform is the current through the* 8 Ω *reference load.*

resistance driven with the same waveform. A ± 10 V output limit is used here for simplicity but this will in practice be higher, a little below the rail voltages.

At the start of the waveform at A, current flows freely into $C_r$ but then reduces to B as the capacitance charges. Current is slowly building up in $L_r$, so the total current drawn increases again to C. A positive transition to the opposite output voltage then takes us to point D, which is not the same as A because energy has been stored in $L_r$ during the long negative period.

A carefully-timed transition is made at E, at the lowest point in this part of the curve. The current change is the same amplitude as at D, but since it starts off from a point where the current is already negative, the final peak goes much lower to 2.96 A, 2.4 times that for the 8 Ω case. I call this the current timing factor (CTF).

## And with multiple speakers?

Otala has shown that the use of multi-way loudspeakers, and more complex electrical models, allows many more degrees of freedom in maximising the peak current, and gives a worst case current timing factor of 6.6 times.[3]

Taking an amplifier designed to give 50 W into 8 Ω, the peak current into an 8 Ω resistance is 3.53 A; amplifiers are usually designed to drive 4 Ω or lower to allow for impedance dips and this doubles the peak current to 7.1 A. In Ref. 3, Otala implies that the peak capability should be at least 23 A, but this need only be delivered for less than a millisecond.

The vital features of the provocative waveform are the fast transitions and their asymmetrical timing, the latter varying with speaker parameters. The waveform in Figure 2 uses ramped transitions lasting 10 μs; as the transitions are made slower the peak currents are reduced. Nothing much changes up to 100 μs, but with 500 μs transitions the current timing factor is reduced from 2.4 to 2.1.

Without doing an exhaustive survey, it is impossible to know how many power amplifiers can supply six times the nominal peak current required. I suspect there are not many. Is this therefore a neglected cause of real audible impairment? I think not, because:

- Music signals do not contain high-level rectangular waveforms, nor trapezoidal approximations to them. A useful step would be to statistically evaluate how often – if ever – waveforms giving significant peak current enhancement occur. As an informal test, I spent some time staring at a digital scope connected to general-purpose rock music, and saw nothing resembling the test waveform of Figure 2. Whether the asymmetrical timings were present is not easy to say; however the large-amplitude vertical edges were definitely not.
- If an amplifier does not have a huge current-peak capability, then the overload protection circuitry will hopefully operate. If this is of a non-latching type that works cleanly, the only result will be rare and very brief periods of clipping distortion when the loudspeaker encounters a particularly unlucky waveform. Such transient distortion is known to be inaudible and this may explain why the current enhancement effect has attracted relatively little attention to date.

## References

1. Self, D. 'Speaker impedance matters', *Electronics World*, Feburary–November 1987, p 920.
2. Otala, M. *et al.* 'Input current requirements of high-quality loudspeaker systems', *AES* preprint #1987 (D7) for 73rd Convention, March 1983.
3. Otala, M. 'Peak current requirement of commercial Huttunen loudspeaker systems', *JAES* June 1987, p 455.
4. Cordell, R. 'Interface intermodulation in amplifiers', *Wireless World*, February 1983, p 32.

# 32 Class distinction

*March 1999*

Some ideas lie fallow for a long time before they see the public light of day, especially if I am involved. This one had been at the back of my mind for some 10 years; the simple thought that complex output stages could be analysed into combinations of a few elementary classes, much as atoms are made of elementary particles.

The goal was not mere classification, but setting up a formal system whereby possible new combinations of operating principles would become evident, and some of these are dealt with – rather briefly, I'm afraid, but space was running out – at the end of the article.

The article was apparently well received at the time, and I received some nice letters, but whether it will have any real impact on the naming of amplifiers remains to be seen. I have to say that my own prediction is no, it will not.

Power amplifiers are usually distinguished by their operating class – A, AB or B, and so on. Unfortunately this classification scheme only begins to address the problem, as amplifiers come in many more than three kinds. There is current-dumping, Class-G, error-correction, and so on. Amplifiers that work in quite different ways are all called 'B' or 'AB', and there is still confusion between B and AB in many quarters.

Traditionally, further letters such as G, H, and S have been used to describe more complex configurations. It occurred to me that rather than proliferating amplifier classes on through the alphabet, it might be better to classify amplifiers as combinations of the most basic classes of device operation.

It may be optimistic to think that this proposal will be adopted overnight, or indeed ever. Nevertheless, it should at least stimulate thought on the many different kinds of power amplifier and the relationships between them.

## Class structure

At the most elementary level, there are five classes of device operation, as outlined in the panel. More sophisticated amplifier types such as Class-G, Glass-S, etc., are combinations of these basic classes. Class-E remains an rf-only technology,[7] while Class-F does not apparently exist.

All the operating classes above work synchronously with the signal. The rare exceptions are amplifiers that have part of their operation driven by the signal envelope rather than the signal itself.

Krell[8] has produced Class-A amplifiers with a quiescent current that is rapidly increased by a sort of noise-gate side-chain, but slowly decays. An interesting study of a syllabic Class-G amplifier with envelope-controlled rail switching was presented in Ref. 9.

## Combinations of classes

The basic classes mentioned in the panel on the right have been combined in many ways to produce the amplifier innovations that have appeared since 1970. Since the standard output stage could hardly be simplified, all of these involve extra power devices that modify how the voltage or current is distributed.

Assuming the output stage is symmetrical about the central output rail, then above and below it there will be at least two output devices connected together, in either a series or parallel format. Since these two devices may operate in different classes, two letters are required for a description, with punctuation – a dot or plus sign – between them to indicate parallel or series connection.

## Parallel or series connection

In parallel, i.e., shunt, connection, output currents are summed, the intention being either to increase power capability, which does not affect basic operation, or to improve linearity.

### Five basic classes

#### *Class-A*

The device conducts 100% of the cycle. This includes Class-A push-pull, where at full output, device current varies from twice the quiescent current to almost zero in a cycle, and Class-A constant-current mode,

also known as single-ended Class-A. Any intermediate amount of current swing clearly also qualifies as Class-A, so unlike Class-B there exists an infinite range of variations on Class-A operation.

### Class-AB

Conducts less than 100% but more than 50% of the cycle. This is essentially over-biased Class-B, giving Class-A operation up to a certain power level, but above that at least twice as much distortion as optimal Class-B. Once more there is a range of variations on Class-AB, depending on the amount of overbias chosen.

### Class-B

The device conducts very nearly 50% of the cycle. The exactness of the 50% depends on the definition of 'conducts' because with Class-B optimally adjusted for minimum crossover distortion, there is always some conduction overlap at crossover, otherwise there would be no quiescent current. This will be 10 mA or so for a complementary feedback pair stage, or about 100 mA for an emitter follower version. With bipolar transistors, collector current tails off exponentially as $V_{be}$ is reduced, and so the conduction period is rather arguable. So-called 'non-switching' Class-B amplifiers, which maintain a small current in the output devices when they would otherwise be off, such as the Blomley[1] and Tanaka approaches[2] are treated as essentially Class-B.

### Class-C

The device conducts less than 50% of the cycle. It is frequently written – indeed I have written it myself – that Class-C is inapplicable to audio and never used therein. A little more thought showed me that this is untrue. The best-known example is Quad current-dumping, a scheme specifically intended to allow the high-power output stage – the 'current-dumpers' – to be run at zero quiescent.[3]

An emitter-follower output stage with no bias has a fixed dead-band of approximately ±1.2 V, so clearly the exact conduction period varies with supply voltage; ±40 V rails and a 1 mA criterion for conduction give 48.5% of the cycle. This looks like a trivial deviation from 50%, but crossover distortion prevents direct audio use.

### Class-D

The device conducts for any percentage of the cycle but is either fully on or off. Class-D usually refers to a pulse-width modulation scheme where the mark/space ratio of an ultrasonic squarewave is modulated by the audio signal.[4,5,6] However, in this case I am concerned only

with the on-off nature of operation, which can be of use at audio frequencies, though not of course for directly driving the load. The conduction period during a cycle is not specified in this definition of Class-D.

A subordinate aim is often the elimination of the Class-B bias adjustment. The basic idea is usually a small high-quality amplifier correcting the output of a larger and less linear amplifier. For a parallel connection the two class letters are separated by a dot, i.e. '•'.

In a series connection the voltage drop between supply rail and output is split up between two or more devices, or voltages are otherwise summed to produce the output signal. Since the collectors or drains of active devices are not very sensitive to voltage, such configurations are usually aimed at reducing overall power dissipation rather than enhancing linearity. Series connection is denoted by a plus sign between the two Class letters.

The order of the letters is significant. The first letter denotes the class of that section of the amplifier that actually controls the output voltage. Such a section must exist – if only because the global negative feedback must be taken from one specific point – and the voltage at this point is the controlled quantity. The shunt configurations are dealt with first; see Table 1.

**Table 1** *Sub-class definitions*

| | | |
|---|---|---|
| *Parallel* | | |
| A • B | Sandman Class-S scheme | Figure 1 |
| A • C | Quad current-dumping | |
| B • B | Self Load-Invariant amplifier | |
| B • C | Crown and Edwin types | Figure 2 |
| B • C | Class-G shunt. (Commutating) 2 rail voltages | Figure 3 |
| B • C • C | Class-G shunt. (Commutating) 3 rail voltages | Figure 4 |
| *Series* | | |
| A + B | 'Super Class-A' | Figure 5 |
| A + B | Stochino error correction | Figure 6 |
| A + D | A possible approach for cooler Class-A | |
| B + B | Totem-pole or cascade output. No extra rails | Figure 7 |
| B + C | Classical series Class-G, 2 rail voltages | Figure 8 |
| B + C + C | Classical series Class-G, 3 rail voltages | |
| B + D | Class-G with outer devices in D | Figure 9 |
| B + D | Class-H | Figure 10 |

### *Class A • B*

Class A • B describes an output stage in which the circuitry that actually controls the output is in Class-A, while a second Class-B stage is connected in parallel to provide the muscle.

The best-known example is probably the Sandman output configuration, in which the high-power amplifier $A_2$ is controlled by its own negative feedback loop so as to increase the effective load impedance until it is high enough for the Class-A stage to drive it with low distortion.[10]

In Figure 1, $A_1$ is the Class-A controlling amplifier while $A_2$ is the Class-B heavyweight stage. As far as the load is concerned, these two stages are delivering current in parallel. The aim was improved linearity, with the elimination of the bias preset of the Class-B stage as a secondary goal.

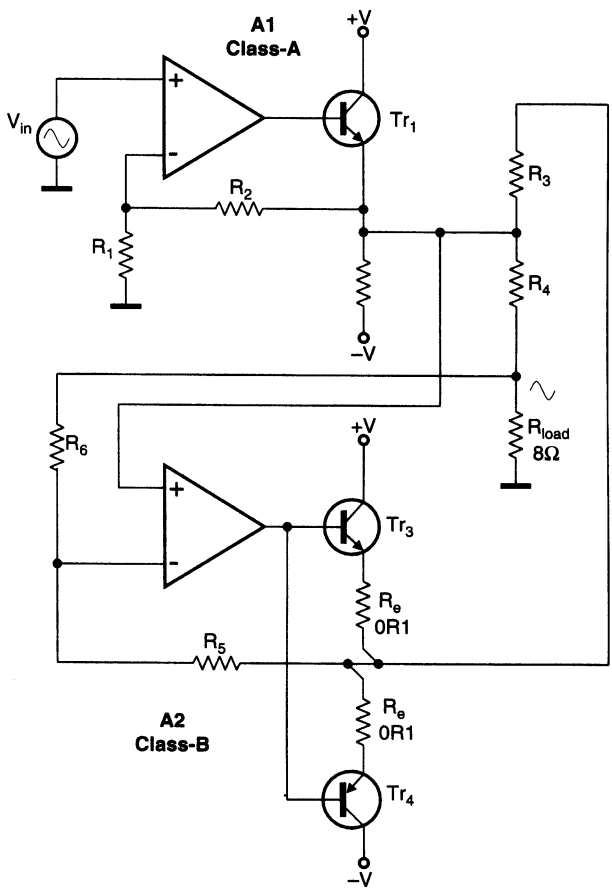



**Figure 1** *Sandman 'Class-S' scheme. Resistors* $R_{3,4,5,6}$ *implement the feedback loop controlling amplifier* $A_2$ *so as to raise the load impedance seen by* $A_2$.

If $A_2$ is unbiased and therefore working in Class-C, $A_1$ has much greater errors to correct. This would put the amplifier into the next category, Class A • C.

### *Class A • C*

The power stage $A_2$ is now working in Class-C, the usual motivation being the reduction of power dissipation because current is flowing for less of the cycle. The absence of any bias for a Class-B-type output stage puts it in into Class-C, as conduction is less than 50% – though probably not much less.

If the bias voltage is dispensed with then a number of problems with setting and maintaining accurate quiescent conditions are eliminated. A good example of such use of Class-C is the Quad current-dumping concept. Here, the use of feed-forward error-correction allows the substantial crossover distortion from a heavyweight Class-C – i.e. underbiased Class-B – stage to be effectively corrected by a much smaller Class-A amplifier.[3]

### *Class B • B*

At first there seems little point in using one Class-B stage to help another, as they both have inherent crossover distortion. However, since reducing the current handled by an output stage reduces both crossover and large-signal distortion, the concept can be useful.

An example is my load-invariant amplifier, which can be considered as two Class-B output stages collapsed into one.[12]

### *Class B • C*

Here, the controlling stage $A_1$ is Class-B, accepting that some crossover distortion in the output will be inevitable. This approach appears to have been introduced by Crown (Amcron) around 1970.[13]

Once more two stages are combined; the drivers – usually compound – are required to deliver significant power in Class-B, while the main power devices only turn on when the output is some way from the crossover point, and are in Class-C.

Similarly, the 'Edwin' type of amplifier, Figure 2, was promoted by Elektor in 1975.[14] It was claimed to have the advantage of zero quiescent current in the main output devices-though why this might be an advantage was not stated; in simulation linearity appears worse than usual.

Another instance of B • C is Class-G-shunt.[11] Figure 3 shows the principle; at low outputs only $Tr_{3,4}$ conduct, delivering power from the low-voltage rails. Above a threshold determined by $V_{bias3}$ and $V_{bias4}$, $D_1$ or $D_2$ conducts and $Tr_{6,8}$ turn on, drawing from the high-voltage rails.

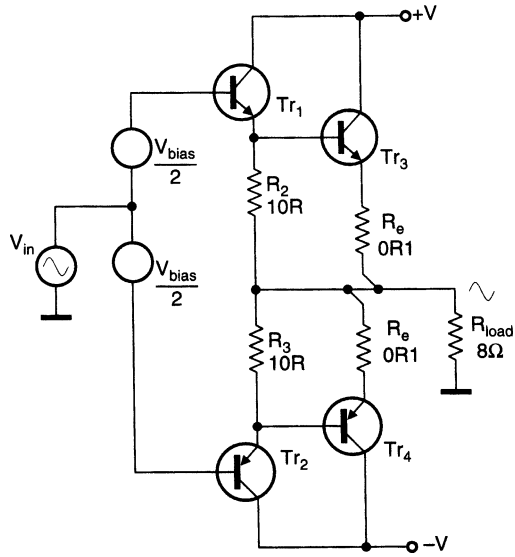



**Figure 2** *Edwin type amplifier; standard Class-B except for the unusually low driver emitter resistors. Effectively B • C.*

Diodes $D_{3,4}$ protect $Tr_{3,4}$ against reverse bias. The conduction periods of the Class-C devices are variable, but much less than 50%. Class-G-Shunt schemes usually have A1 running in Class-B to minimise dissipation, giving B • C; such arrangements are often called 'commutating amplifiers'.

### *Class B • C • C*

Some of the more powerful Class-G-shunt public-address amplifiers have three sets of supply rails to further reduce the average voltage-drop between rail and output.

The extra complexity is significant, as there are now six supply rails and at least six power devices. It seems most unlikely that this further reduction in power consumption could ever be worthwhile for domestic hifi, but it is very useful in large PA amplifiers, such as those made by BSS. Three letters with intervening dots are required to denote this mode, Figure 4.

## Series connection category

In the second group of configurations, voltages are summed by series connection. The intention is usually the reduction of total power dissipation, rather than better linearity.

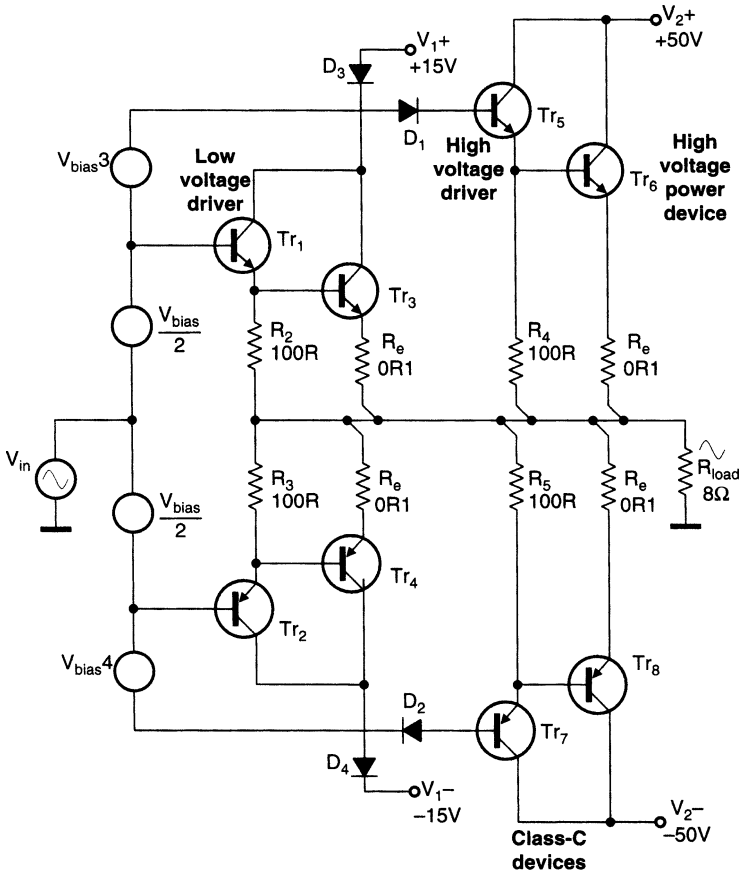



**Figure 3** *A Class-G-shunt output stage, composed of two emitter-follower output stages with the usual drivers. Voltages* $V_{bias3,4}$ *set the output level at which power is drawn from the higher rails. B • C.*

Since the devices are not usually operating in the same class, two letters are again required for a description, and I have used a plus-sign between them to indicate the series connection.

## *Class A + B*

Figure 5 shows the so-called 'Super-Class-A' introduced by Technics in 1978.[15] The intention is to combine the linearity of Class-A with the efficiency of Class-B.

The Class-A controlling section $A_1$ is powered by two floating supplies of relatively low voltage, around ±15 V, but handles the full load current.

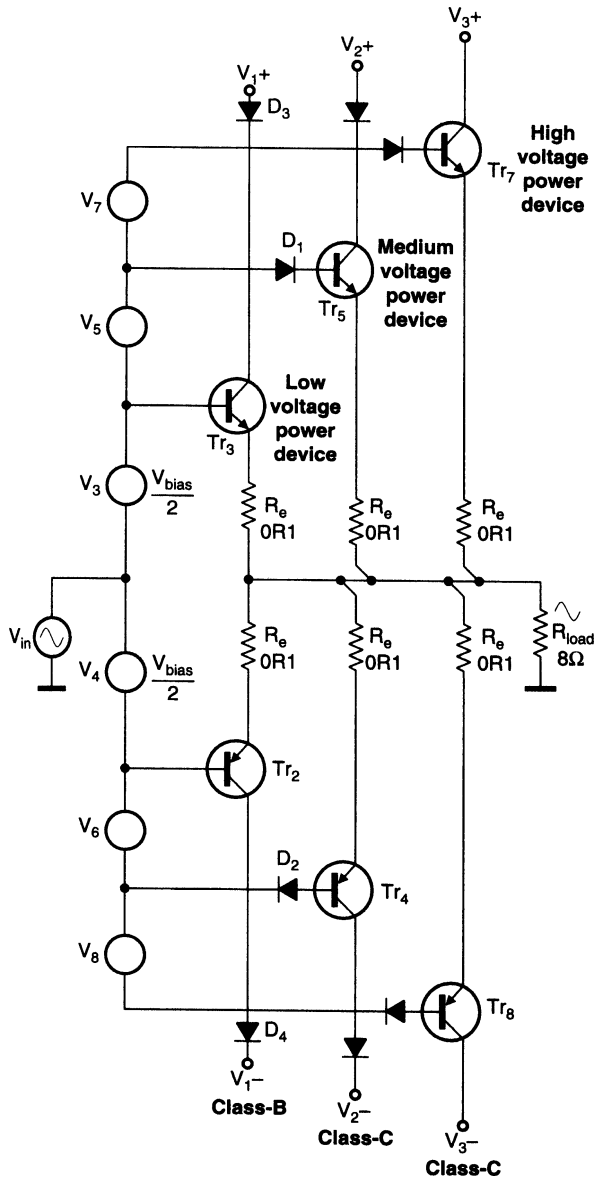



**Figure 4** *Simplified diagram of a three-rail 'commutating' series-Class-G power amplifier, denoted as Class B • C • C.*

The floating supplies are driven up and down by a Class-B amplifier $A_2$. This amplifier must sustain much more dissipation as the same current is drawn from much higher rails, but it need not be very linear as in principle its distortion will have no effect on the output of $A_1$.

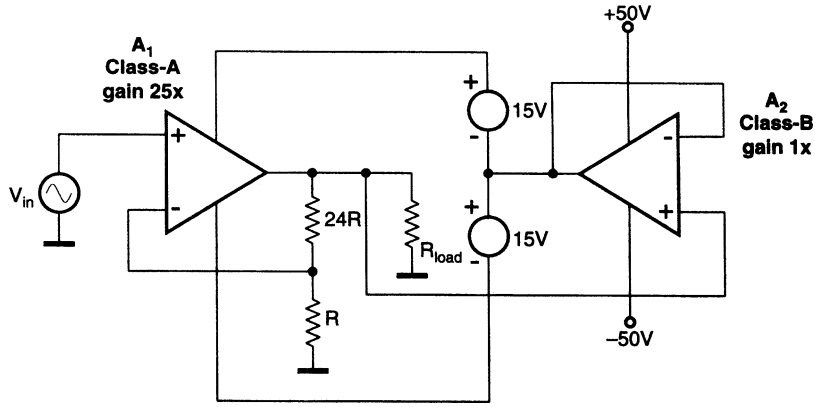



**Figure 5** *The 'Super-Class-A' concept. Amplifier $A_1$ runs in Class-A, while high-power Class-B amp $A_2$ drives the two floating supplies up and down. Denoted as A + B.*

The circuit is complex and costs more than twice that of a conventional amplifier. In addition, the floating supplies are awkward. This seems to have limited its popularity.

Another A + B concept is the error-correction system of Stochino.[16] The voltage summation – the difficult bit – can be performed by a small transformer, as only the flux due to the correction signal exists in the core. This flux cancellation is enforced by the correcting amplifier feedback loop. Complexity and cost are at least twice that of a normal amplifier; Figure 6.

## *Class A + D*

The 'Super-Class-A' concept mentioned above can be extended to A + D by running the heavyweight amplifier in the usual high-frequency pwm Class-D configuration.[17] Alternatively, an A + D amplifier can be made by retaining the Class-A stage but powering it from rails that switch at audio frequency between two discrete voltages. Recall that this definition of Class-D does not mean high-frequency pwm.

## *Class B + B*

Sometimes called a totem-pole stage to emphasise the vertical stacking of output devices, this arrangement shares the power dissipation between two devices. However, a parallel connection does the same thing more simply and with lower voltage losses.

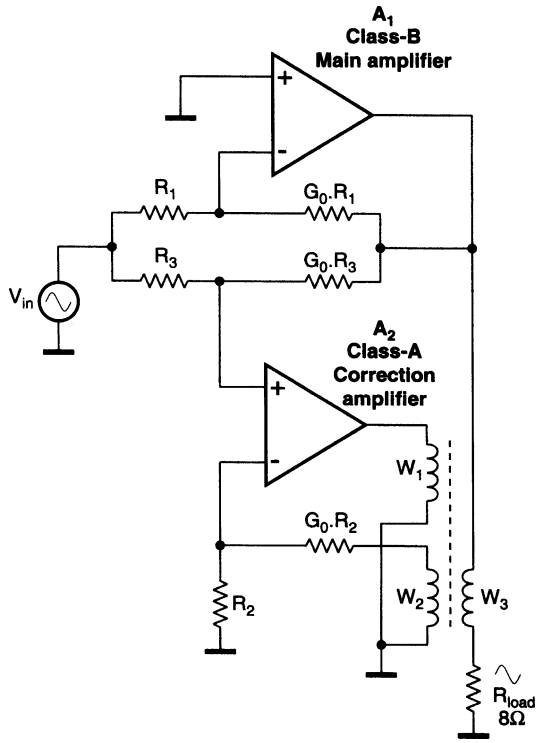



**Figure 6** *The Stochino error-correcting system voltage-sums the outputs of the two amplifiers using transformer* $W_{1,2,3}$, $A+B$.

Class B + B has been used to permit high power outputs from transistors with limited $V_{ceo}$, but this is rarely necessary with modern devices. The concept is usually regarded as obsolete, Figure 7.

## *Class B + C*

The basic series Class-G with two rail voltages, i.e., four supply rails, as both voltages are positive and negative, is shown in Figure 8. This configuration was introduced by Hitachi in 1976 with the aim of reducing amplifier power dissipation.[18,19]

Musical signals spend most of their time at low levels, and have a high peak/mean ratio, so dissipation is greatly reduced by running from the lower $\pm V_1$ supply rails when possible.

When the instantaneous signal level exceeds $\pm V_1$, $Tr_6$ conducts and $D_3$ turns off, so the output current is now being drawn from the higher $\pm V_2$ rails, with the dissipation shared between $Tr_3$ and $Tr_6$. The inner stage $Tr_{3,4}$

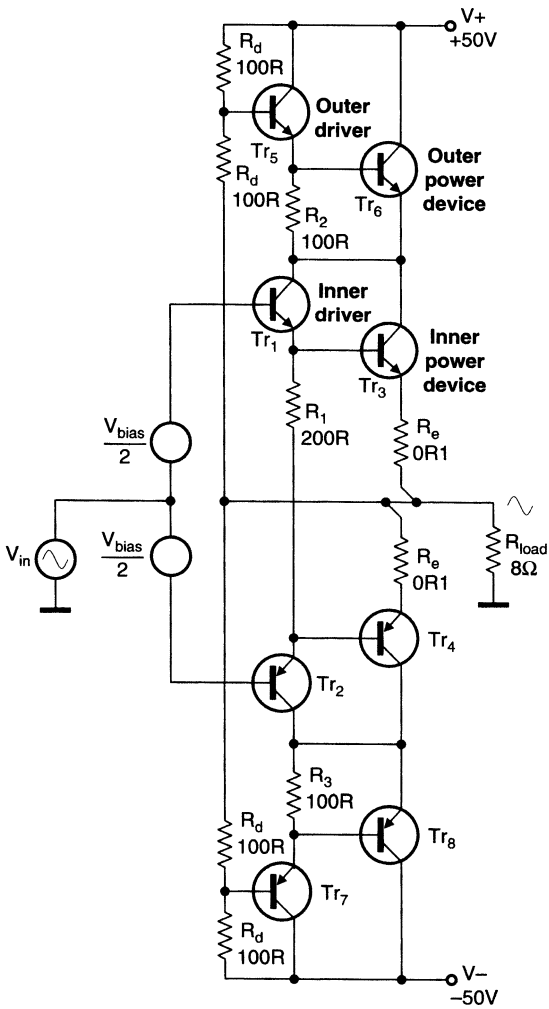



**Figure 7** *A totem-pole or cascade series output. Resistors* $R_d$ *divide the voltage between rail and output in half, and drive the outer power devices. Inner and outer devices turn on and off together. B + B.*

normally operates in Class-B, though AB or A are equally possible if the output stage bias is increased.

In principle movements of the collector voltage on the inner device collectors should not affect the output voltage, but practical Class-G is often considered to have worse linearity than Class-B because of glitching due to diode commutation. However, glitches if present occur at moderate power, well away from the crossover region.

Class B + C + C An obvious extension of the Class-G principle is to increase the number of supply voltages, typically to three. Dissipation is

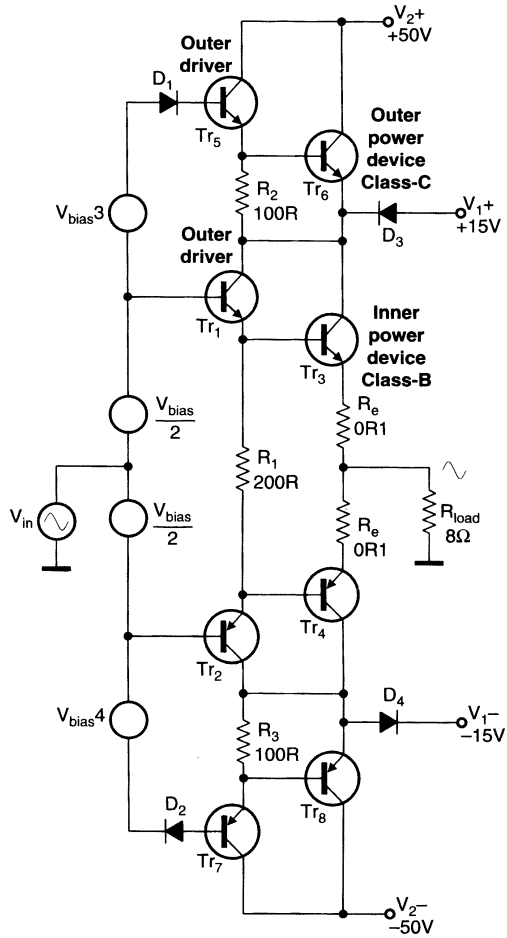



**Figure 8** *Class-G-series output stage. When the output voltage exceeds the transition level,* $D_3$ *or* $D_4$ *turns off and power is drawn from the higher rails through the outer power devices. B + C.*

reduced and efficiency increased, as the average voltage from which the output current is drawn is kept closer to the minimum.

The inner devices will operate in Class B/AB as before, the middle devices will be in Class-C, conducting for significantly less than 50% of the time. The outer devices are also in Class-C, conducting for even less of the time. Three letters with intervening plus signs are required to denote this.

To the best of my knowledge three-level Class-G amplifiers have only been made in shunt mode. This is probably because in series mode the cumulative voltage drops become too great. If it exists, such an amplifier would be described as operating in B + C + C.

### *Class B + D*

Since the outer power devices in a Class-G-series amplifier are not directly connected to the load, they need not be driven with waveforms that mimic the output signal. In fact, they can be banged hard on and off so long as they are always on when the output voltage is about to hit the lower supply rail.

The outer devices may be simply driven by comparators, rather than via a nest of extra bias generators as in Figure 8. Thus the inner devices are in B with the outer in D. Some of the more powerful amplifiers made by NAD – like the *Model 340* – use this approach, shown in Figure 9.

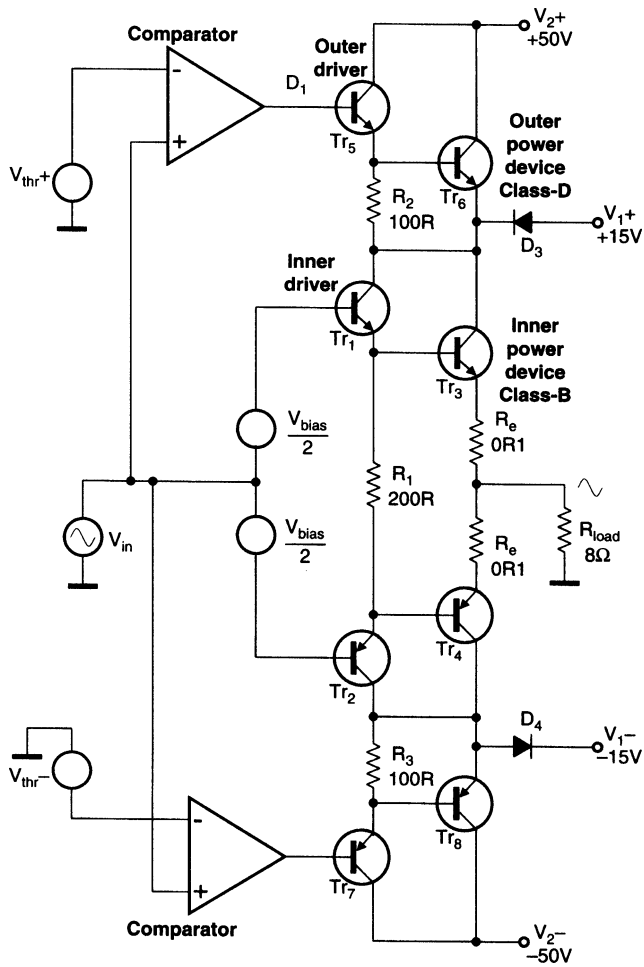



**Figure 9** *Class-G output stage with outer devices in Class-D. Described as B + D.*

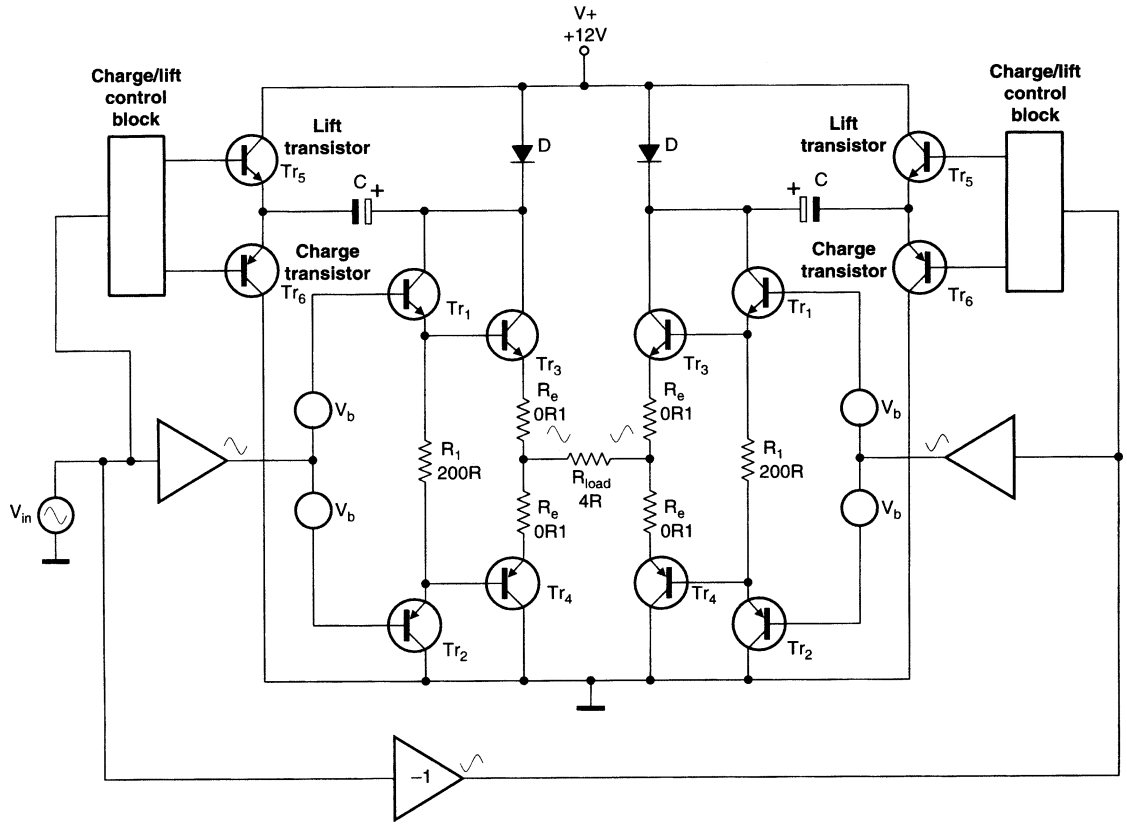



**Figure 10** *The Class-H principle applied to a bridged output stage for automotive use. B + D.*

The technique known as Class-H is similar but uses a charge-pump for short-term boosting of the supply voltage. In Figure 10, at low outputs $Tr_6$ is on, keeping $C$ charged from the rail via $D$.

During large output excursions, $Tr_6$ is off and $Tr_5$ turns on, boosting the supply to $Tr_3$. The only known implementation is by Philips[20,21] which is a single-rail car audio system that requires a bridged configuration and some clever floating-feed-back to function.

Full circuitry has not been released, but it appears the charge-pump is an on/off subsystem, i.e. Class-D.

## In summary

The test of any classification system is its gaps. When the periodic table of elements was evolved, the obvious gaps spurred the discovery of new elements; convincing proof the table was valid.

Table 1 is restricted to combinations that are, or were, in actual use, but a full matrix showing all the possibilities has several intriguing gaps; some, such as C • C and C + C are of no obvious use, but others like A + C are more promising – a form of Class-G with a push-pull Class-A inner stage. Glitches permitting, this might save a lot of heat.

The amplifier table really gets interesting when it becomes clear that there are gaps in the entries – things that could exist but are not currently known.

## References

1. Blomley, P. 'A new approach to class-B', *Wireless World*, February 1971, p 57.
2. Tanaka, S. 'A new biasing circuit for class-B operation', *Journal of Audio Engineering Society,* January/February 1981, p 27 (Non-switching Class-B).
3. Walker, PJ. 'Current dumping audio amplifier', *Wireless World*, December 1975, p 560.
4. Attwood, B. 'Design parameters important for the optimisation of PWM (Class-D) amplifiers', *JAES,* Vol. 31 (11), November 1983, p 842.
5. Goldberg and Sandler, 'Noise shaping and pulse-width modulation for an all-digital audio power amplifier', *JAES*, Vol. 39 (6), February 1991, p 449.
6. Hancock, J. 'A class-D amplifier using MOSFETS with reduced minority carrier lifetime', *Journal of Audio Engineering Society,* Vol. 39 (9), September 1991, p 650.
7. Peters, A. 'Class-E RF amplifiers', *IEEE Journal of Solid-State Circuits*, June 1975, p 168.

8. Atkinson, J. 'Krell KSA-50S amplifier review', *Stereophile*, August 1995, p 165.
9. Funada and Akiya, 'A study of high-efficiency audio power amplifiers using a voltage switching method', *Journal of Audio Engineering Society,* Vol. 32 (10), October 1984, p 755 (Syllabic Class-G).
10. Sandman, A. 'Class-S: A novel approach to amplifier distortion', *Wireless World*, September 1982, p 38.
11. Raab, F. 'Average efficiency of class-G amplifiers', *IEEE Trans on Consumer Electronics*, April 1986, p 145.
12. Self, D. 'Load invariant audio power', *Electronics World*, January 1997, p 16.
13. Linsley-Hood, J. 'The straight wire with gain?', *Studio Sound*, April 1975, p 22. (Crown).
14. Unknown, 'Edwin Amplifier', *Elektor*, September 1975, p 910.
15. Sano *et al.* 'A high-efficiency class-A audio amplifier', preprint #1382 for 61st AES convention, November 1978.
16. Stochino, G. 'Audio design leaps forward?', *Electronics World*, October 1994, p 818. (Error correction).
17. Jeong *et al.* 'A high-efficiency Class-A amplifier with variable power supply', (ClassA+Class-D), preprint #4257 for 100th AES convention, May 1996.
18. Sampei, *et al.* 'Highest efficiency & super quality audio amplifier using MOS power FETs in Class-G operation', *IEEE Trans on Consumer Electronics*, Vol. CE-24, #3 August 1978, p 300. (Class-G).
19. Feldman, 'Class-G high efficiency Hi-Fi amplifier', *Radio-Electronics*, August 1976, p 47.
20. Buitendijk, P. 'A 40 W integrated car radio audio amplifier', *IEEE Conf on Consumer Electronics*, 1991 Session THAM 12.4, p 174. (Class-H).
21. Philips, 'TDA1560Q 40 W car radio high power amplifier', Philips Semiconductors' data sheet, April 1993.

# 33 Muting relays

*July 1999*

What could be simpler than switching a relay on and off? This depends on how closely you want to control it. Any design problem, when looked at the right way, can be made more challenging. It is usually possible to identify a parameter that can be focused on and made the subject of some creative improvement. So it is here, where the speed of relay switch-off, and hence its ability to mute turn-off transients, is examined into in depth. This article gives amongst other advice, details on how to make relays operate faster than at first appears to be possible.

Since this article was published, further experience leads me to sound a note of warning about trying to be too quick in sensing the disappearance of the mains supply. Doubtful electrical appliances (and I am thinking of a particular veteran refrigerator here, with a vicious motor start-up transient) combined with antique mains wiring can put short-duration dips in the supply that trigger the AC-loss circuit, when in fact the hold-up time of the power supply in the equipment with the relay is perfectly capable of keeping the show going. If you run into trouble with this, increasing the value of Cl (Figure 3) should fix things.

Most power amplifiers incorporate an output relay that not only provides muting to prevent transients reaching the loudspeakers, but also protection against destructive DC faults.

Loudspeakers are expensive, and no amplifier should ever be connected to one without proper DC-offset protection. This applies with particular force to experimental amplifiers.

Sensible preamplifiers – i.e. those with AC-coupled outputs – do not require DC protection, but the muting of thumps is no less important. Electronic switching at preamp outputs is feasible, but still presents technical

challenges if high standards of linearity are to be combined with a reasonably low output impedance.

Electronic output switching is impracticable at power amplifier signal levels; however, if the amplifier is powered by a switch-mode supply, then turning it off is an option if positive and negative rails can be relied upon to collapse quickly and symmetrically.

## Protection circuit operation

Basic functions of a power-on thump elimination and DC protection circuit are as follows:

- Delay relay pull-in until amplifier turn-on transients are over.
- Drop out relay as fast as possible when AC power is removed.
- Drop out relay as fast as possible when DC fault occurs.
- Drop out relay on excess temperature, etc. Speed non-critical.

Figure 1 is a block-diagram of a system to perform these functions. Since this is in part a protection system, simplicity and bullet-proof reliability are essential.

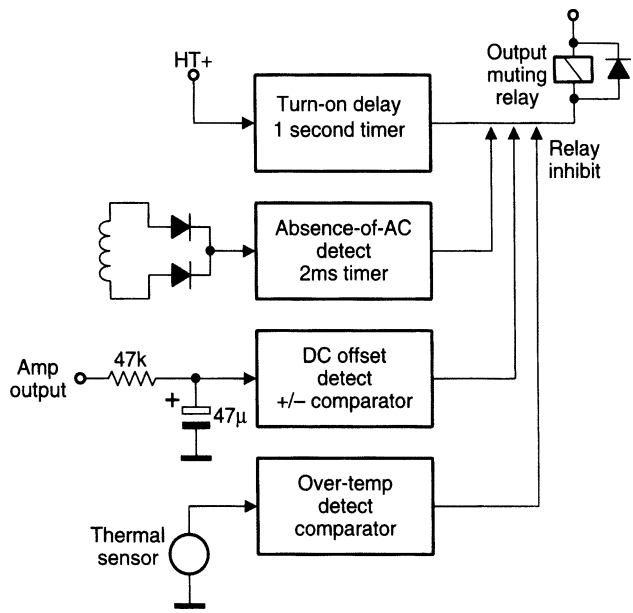



**Figure 1** *Block diagram of a relay control system.*

The main dynamic parameters of a relay are the pull-in and drop-out time. For this kind of application, the pull-in time is more or less irrelevant, as it is milliseconds compared with the seconds of the turn-on delay.

Relay contacts bounce when they close, but the duration of pull-in contact bounce is not important for this application.

All the relays I examined showed clean contact-breaking on drop-out, and this is essential for fast muting. Table 1a gives details of three power-amp relays and the Fujitsu relay used in the Precision Preamp '96 article.[1]

The specifications for the P&B relay are very conservative. The example measured pulled-in at 72% of the must-operate voltage, and dropped out at 350% of the must-drop-out voltage. Likewise the real operating times are much less than those specified.

The critical parameter for audio muting is the drop-out time, for this puts a limit on the speed with which turn-off transients can be suppressed. It seems at first that the drop-out time must be solely a function of the relay design, depending on the force in the bent contact spring and the inertia of the moving parts. This is partly true, as mechanical factors set

**Table 1a** *Relay specifications as presented by their manufacturers*

| | *P&B* | *Oko* | *Schrack* | *Fujitsu* |
|---|---|---|---|---|
| Nominal voltage | 24 V | 18 V | 12 V | 12 V |
| Must-operate voltage | 18 V | | | 8.4 V |
| Drop-out voltage | 2.4 V | | | 1.2 V |
| Coil resistance | 660 Ω | | | 320 Ω |
| Coil inductance | 0.55 H | | | |
| Pull-in time maximum | 15 ms | | | 5 ms |
| Pull-in time typical | 9 ms | | | |
| Drop-out time maximum | 10 ms | | | 3 ms |
| Drop-out time typical | 7 ms | | | |

**Table 1b** *Measured relay specifications*

| | *P&B* | *Oko* | *Schrack* | *Fujitsu* |
|---|---|---|---|---|
| Operate voltage | 13 V | 13 V | 7 V | 6 V |
| Drop-out voltage | 8.5 V | 6.5 V | 2.5 V | 2 V |
| Pull-in time | 14 ms | 10 ms | 10 ms | 2.7 ms |
| Drop-out time | 1.0 ms | 1.3 ms | 2.4 ms | 1.2 ms |
| Diode drop-out time | 5.4 ms | 6.9 ms | 11 ms | 4.2 ms |
| 27 V-clamp drop-out time | 1.8 ms | 2.4 ms | 2.7 ms | 1.3 ms |

* P&B is Potter and Brumfield.

a minimum time, but that time is greatly extended by the normal relay-driving circuits.

## Relay-on timing

The delay required at amplifier turn-on depends on the amplifier characteristics.

If there are long time-constants, and voltages that take a while to settle, then the muting period will have to be extended to prevent clicks and thumps. Five seconds is probably the upper limit before the delay gets irritating; one second is long enough for a silent start-up with most conventional amplifiers.

This delay function can be performed in many ways, but there are a few points to consider. The tolerance on the length of the turn-on delay is not critical, and an *RC* time-constant is quite adequate to define it.

It is convenient – and significantly cheaper – to run the relay control circuitry directly from the main HT rails rather than creating regulated sub-rails or extra windings on the mains transformer. The emphasis is therefore on discrete transistor circuitry.

Figure 2 shows the relay control system I used in the Precision Preamplifier '96. Note that there was an error in the original diagram that is corrected here.

Capacitor $C_{224}$ charges through $R_{211}$ until $D_{207}$ is forward-biased and $Tr_{205}$ turns on. This turns on $Tr_{206}$ and energises the relay; the extra current-gain of $Tr_{206}$ enables the timing circuitry to run at low power. The on-timing delay here is 2 s.

A series dropper resistor for the relay is usually required; here it is $R_{218}$. The highest voltage relay-coil available is usually 48 V, though 24 V is more common, and power amplifier rails are often much higher than this.

This reverse diodes across the relay coils prevent $Tr_{206}$ being damaged by the inductive spike created when the coil is suddenly de-energised. For relays of the size used in power amplifiers, signal diodes cannot cope with the stored energy and the 1N4001 type should be used.

## Off timing criteria

The relay drop-out must be as fast as possible. If a relay is powered directly from the supply rails, then it will drop out eventually as the rails collapse, but this will be far too slow to catch turn-off noises.

The drop-out voltage may well be less than a third of the pull-in voltage, and this slows things down even more. A specific fast off circuit is required, and there are several ways to achieve this.

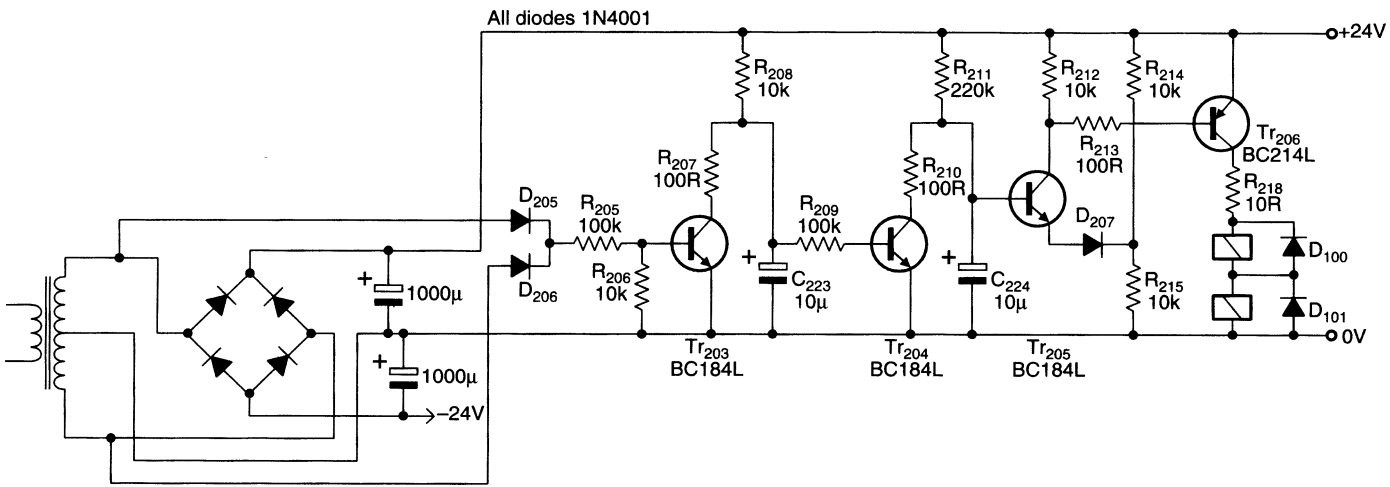



**Figure 2** *Relay-control circuit as used in the Precision Preamplifier '96 – effective, but it can be improved. Component numbers retained from original design.*

### *Mechanical detection*

This is a mains switch that closes or opens a control circuit before the mains power contacts are opened. It could give perfect relay operation, but I am not aware that any such mains switch has ever been produced.

### *Detecting the loss of DC supply*

This technique involves a subsidiary supply rail with a small reservoir capacitor. When the mains is switched off, the capacitor discharges quickly and either removes the relay power directly, or resets the turn-on delay timer. The latter is usually easier to implement.

This method is inherently slow, because the relay-off threshold must be below the ripple troughs. Therefore in the worst case, an entire half-cycle of mains must pass before the capacitor becomes fully discharged, so the delay may be 10 ms.

In practice there are component tolerances to be allowed for, and the threshold must be set low enough to prevent spurious operation if the mains voltage is below normal. It is usually prudent to ensure circuitry works with mains down to at least −20%. This extends the minimum delay to about 16 ms. The reservoir capacitor will have a large ripple voltage across it, and its ripple-current rating must be carefully observed.

### *Detection of loss of AC*

Detecting the loss of AC supply, as opposed to the rectified DC, is potentially quicker as there is no reservoir capacitor to discharge before the circuit operates. An AC waveform is effectively appearing and disappearing every half-cycle, so the circuit must distinguish between the zero-crossings that occur every 10 ms, and genuine loss of power.

## AC loss detection

The most straightforward method of AC-loss detection exploits the fact that properly-defined zero-crossings are very brief; all that is required is a timer that will not complete and drop out the relay until a period greater than the width of the zero-crossing has expired. This delay can readily be reduced to less than 1 ms.

Referring to Figure 2, $Tr_{203}$ is normally held firmly on by the incoming AC, via $D_{205}$ on positive half-cycles and $D_{206}$ on negative ones, and thus keeps $C_{223}$ fully discharged.

At the zero-crossings, $Tr_{203}$ has no base drive and turns off, allowing $C_{223}$ to begin charging through $R_{208}$. If the absence of base drive persists beyond the preset period, which means the AC has been interrupted, then

$C_{223}$ charges until $Tr_{204}$ turns on and rapidly discharges the main timing capacitor $C_{210}$ through $R_{207}$, dropping out the relay.

When a relay is driven by a transistor, it is standard to put a reversed diode across the coil. Without it, abrupt turn-off of current causes the coil voltage to reverse, driving the collector more negative. For the relays here, the worst spike measured was −120 V, which is enough to destructively exceed the $V_{ceo}$ of most transistors.

This apparently innocent, and indeed laudable practice of diode protection conceals a lurking snag; drop-out time is hugely increased by the reversed diode. It is roughly five times longer, which is very unwelcome in this particular application. This is because the diode gives a path for current to circulate while the magnetic field decays.

This is a good point to stop and consider exactly what we are trying to do: the aim is not to totally suppress the back-EMF but rather to protect the transistor.

If the back-EMF is clamped to about −27 V by a suitable Zener diode in series with the reverse diode, the circulating current stops much sooner, and the drop-out is almost as fast as for the non-suppressed relay.

In general, drop-out is speeded up by a factor of about four on moving from conventional protection to Zener clamping. For the relays examined here, a 500 mW Zener appeared to be adequate.

## Preamp enhancement

The preamp relay controller can be improved upon; it works well under most circumstances, but it could be faster. Testing showed that the delay between loss of AC and the relay power being removed could be as long as 17 ms, depending slightly on the phase of the mains when it was cut. The relay drop-out time was 5 ms giving a total of 22 ms before the preamp output is muted.

The following circuit improvements were made to speed up relay drop out.

The on-timing reference divider $R_{214,215}$ is replaced with a 15 V Zener diode. This sharpens up the relay pull-in, making a more 'precise' click. It also prevents the voltage on $C_{224}$ rising beyond that required to turn on $Tr_{205}$; discharging it when the time comes is therefore quicker.

Base drive to $Tr_{203}$ is increased by reducing $R_{205}$ to 22 kΩ. This defines the zero-crossing as twice as narrow, allowing the time-constant $R_{208}$–$C_{223}$ which bridges this period to be made shorter. Capacitor $C_{224}$ therefore starts discharging sooner after AC is lost.

Impedance of the zero-crossing time-constant $R_{208}$–$C_{223}$ is increased by changing the values from 10 kΩ–10 μF to 100 kΩ–470 nF. This simultaneously reduces the time-constant mentioned in the previous paragraph. It

**Table 2** *Component revisions for the preamplifier*

| | *Old value* | *New value* |
|---|---|---|
| $R_{205}$ | 100 kΩ | 22 kΩ |
| $R_{208}$ | 10 kΩ | 100 kΩ |
| $C_{223}$ | 10 µF | 470 nF |
| $R_{209}$ | 100 kΩ | 100 Ω |
| $R_{215}$ | 10 kΩ | 15 V 400 mW Zener |
| Relay | 1N4001 | 1N4001 + Suppression 27 V, 500 mW Zener |

is now possible to use a non-electrolytic timing capacitor, which reduces tolerances and makes the circuit more designable.

Base drive to $Tr_{204}$ is increased to speed up the discharge of $C_{224}$ by reducing $R_{205}$ from 100 kΩ to 100 Ω.

Finally, a 27 V Zener clamp is applied to each relay, as described above.

After these improvements, the electronic delay was reduced from 17 to 5.4 ms; the total delay including contacts opening now was 9.5 ms worst case.

After adding Zener clamping to the relays this fell to 6.3 ms worst-case, the average being 4.5 ms; the improved circuit is four times faster.

These component changes can be simply retro-fitted to existing Preamp 96 circuit boards using Table 2.

## Other relay functions

The extra protective functions of a power amp relay require OR-ing together several error signals for DC offset, temperature shutdown, etc.

If a DC fault occurs in a power amplifier, this typically means that the output slams hard to one of the rails and stays there. Assuming the loudspeaker does not suffer instantaneous mechanical damage, it will overheat after a relatively short period as the DC flows through it.

DC offset protection cannot prevent a loudspeaker hitting its mechanical limits, but it will stop it catching fire if the relay opens promptly. Once more, time is of the essence.

Usually, DC offset is detected by passing the amplifier output through an *RC* time-constant long enough to remove all audio, followed by a DC-detect circuit that responds to offsets of either polarity.

To allow a safety margin against false triggering on bass signals, I decided that the *RC* filter must accept full output at 2 Hz without the detector acting. For example, if it triggers at ±2 V, then for supply rails of ±55 V there must be 29 dB of attenuation at 2 Hz; with a single pole this means a −3 db frequency about 0.07 Hz.

This sort of low-pass filtering inevitably introduces a time delay; if the output leaves 0 V and moves promptly to one of the rails, this will be 50 ms with the circuit of Figure 3.

Detecting offsets of either polarity requires a little thought. Figure 4 shows a common circuit; a positive voltage turns on $Tr_8$ by forward biasing its base, while a negative voltage turns on $Tr_7$ by pulling down the emitter. The presence of DC is indicated by the collector voltage falling.

This solution is simple but highly-asymmetrical, requiring either +1.05 V or −5.5 V to pull the collectors down to 0 V. For positive voltages the stage is common-emitter with high voltage gain, but for negative ones it works in common-base with a lower voltage gain, set by the ratio of $R_{16}$ and $R_9$. It is difficult to make this ratio large without $R_{16}$ becoming too small and hence $C_4$ inconveniently big.

If you're unlucky – and chances are you will be – the offset will have the wrong polarity for $C_4$, which will degrade if left reverse-biased for long periods. Two ordinary electrolytics back-to-back is the cheapest solution.

The improved DC detector in Figure 5 is fully symmetrical. Positive voltages turn on $D_{11}$ and $Tr_8$; $Tr_7$ also conducts as its emitter is pulled up by $Tr_8$, while its base is held low by $D_{14}$. Negative voltages turn on $D_{13}$ and $Tr_7$; $Tr_8$ conducts with its base fed by $D_{12}$. The threshold is now ±2.4 V, as for each polarity there are two diodes and two base-emitter voltages in series. The higher threshold is not a problem as the typical amplifier fault snaps the output hard to one of the rails.

One exception to this statement is HF instability. If an amplifier bursts joyfully into HF oscillation, it will almost certainly show slew-limiting as well. This is unlikely to be very symmetrical so there will be a DC shift at the output.

The magnitude of this is not very predictable, but a 2.4 V threshold will detect most cases. This should save your tweeters, though it may not save the amplifier from internal heating due to conduction overlap in the relatively slow output devices.

Figure 5 can be adapted for stereo simply by adding two more diodes, as in Figure 3. Note that a positive offset on one channel and a negative one on the other – admittedly highly unlikely – do not cancel out; a fault is still signalled.

## Power amplifier relay control

Figure 3 shows a power amp relay controller, designed for ±55 V rails and 24 V relays such as the P&B *T*90 type outlined in Table 1a. The main differences are the inclusion of DC offset detection and an efficiency circuit to minimise dissipation in the relays, which are now larger than in the preamp, and require more power.

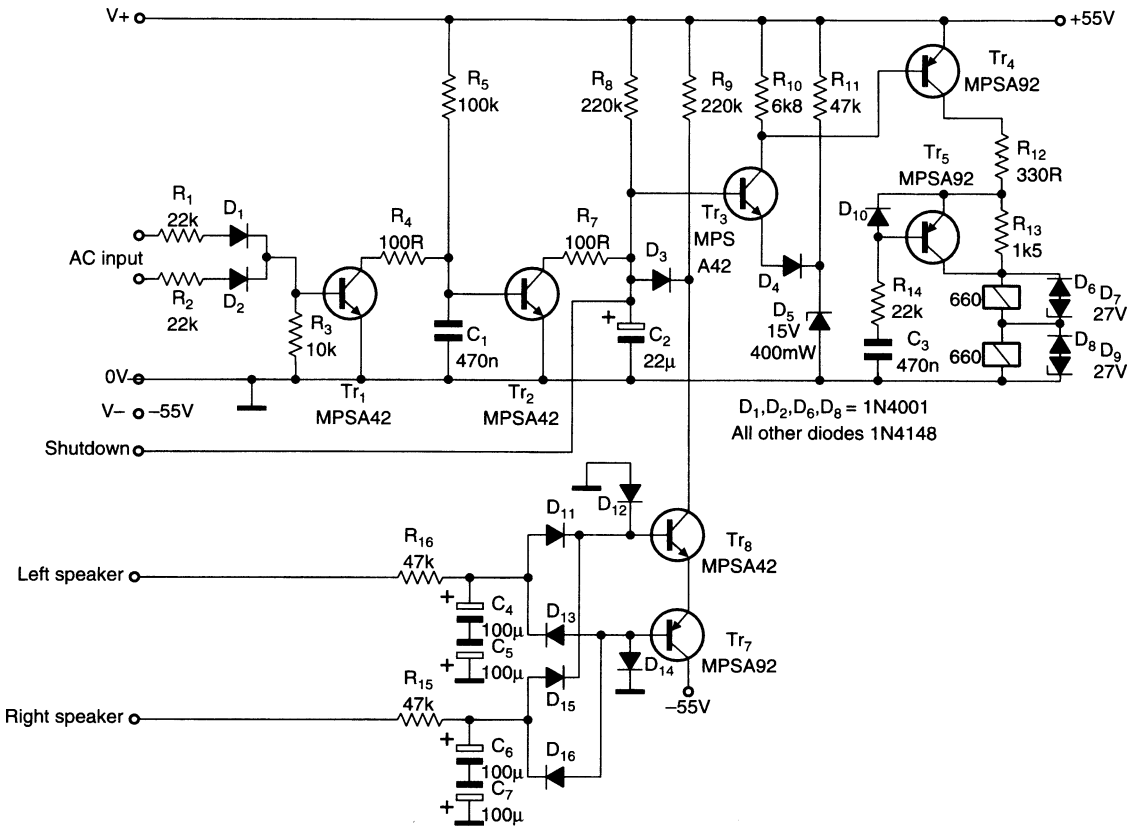



**Figure 3** *Relay control circuit for power amplifier, incorporating the efficiency circuit* $Tr_5$.

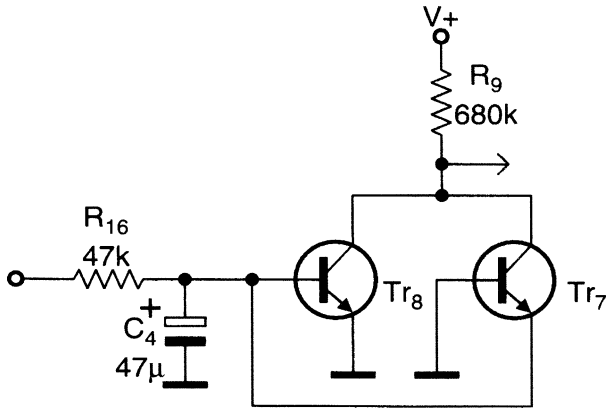



**Figure 4** *Simple DC-detect circuit with asymmetrical thresholds at* +1.05 V *and* −5.5 V.

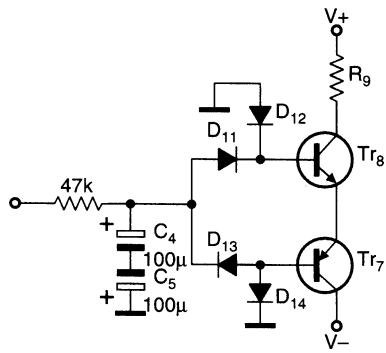



**Figure 5** *Improved DC-detect circuit; fully symmetrical thresholds at* ±2.4 V. *RC filter can cope with either polarity.*

The DC-detect circuit rapidly discharges on-timing capacitor $C_2$ through $D_3$ when $Tr_8$ collector goes low. An extra OR input for thermal shutdown acts via a series diode in the shutdown line.

The circuit now uses *MPSA42/MPSA92* transistors to withstand the higher supply voltages; as usual higher $V_{ceo}$ means lower current-gain, which must be allowed for in the detailed design.

The electronic delay until coil switch-off averages 2 ms, the timing being shown in Figure 6. The AC was interrupted at centre screen, and a large positive-going off-transient can be seen just to the right. This is due to the leakage inductance of a large transformer.

The loss of AC cannot be detected until this transient decays to zero, so the delay is slightly extended. This was not a problem with the preamp version as it uses a small toroid with much less leakage inductance.

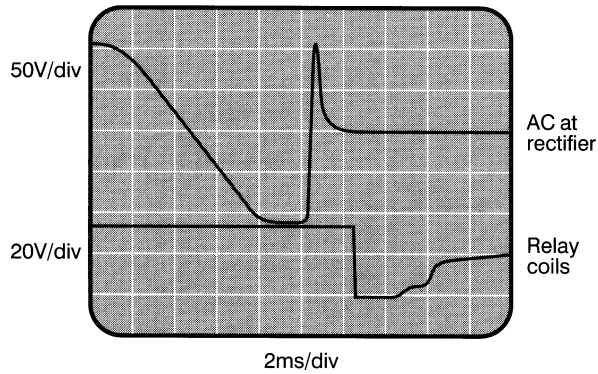



**Figure 6** *Electronic delay for power-amp version. Upper trace is transformer secondary with the usual flat-topped mains waveform, lower is voltage across both relay coils.*

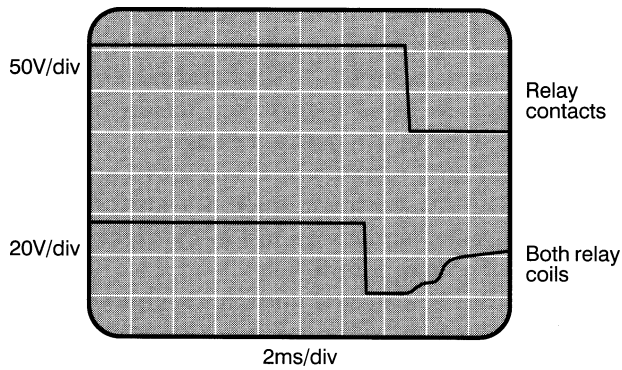



**Figure 7** *Relay drop-out delay for power-amp version. Upper trace is contact timing, lower is voltage across both relay coils.*

Figure 7 shows the relay coil voltage. At switch off it goes straight down through zero until clamped by the two Zeners at around −50 V. This puts 105 V on $Tr_4$ collector, which is no problem as it is rated at 300 V $V_{ceo}$. The relay contacts open just as clamping ceases and the coil voltage returns slowly to zero. Drop-out time is 1.8 ms, giving a total delay of 3.8 ms.

## Efficiency circuit

All relays have a pull-in voltage that is greatly in excess of that required to keep them closed. It is therefore possible to save considerable power by applying full voltage only briefly, and then reducing it to a level which is still safely above the maximum drop-out voltage.

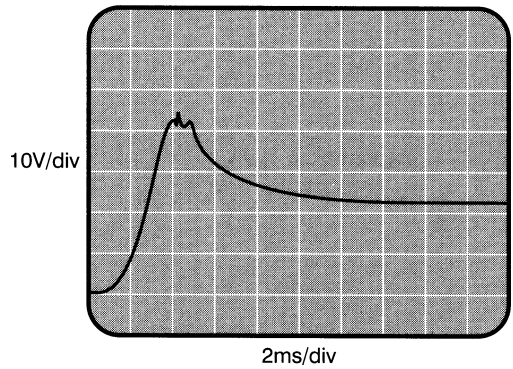



**Figure 8** *Pull-in voltage across both relay coils with efficiency circuit added.*

From Table 1a there is plenty of scope for this. By comparing the specified and measured performance, you will see that the P&B relay can be trusted to pull in at 18 V and not drop out above 8.5 V.

The initial pulse is provided by $Tr_5$ and $R_{13}$. At switch-on, $Tr_4$ is off and $Tr_5$ does nothing. After the on-timing delay $Tr_4$ conducts, $Tr_5$'s emitter is pulled up, and its base receives a pulse of current via $R_{14}$ and $C_3$. Resistor $R_{13}$ is shorted by $Tr_5$ and the relays get a voltage reduced only by $R_{12}$; see Figure 8.

After 40 ms, $C_3$ is fully charged and $Tr_5$ turns off; this is at least four times longer than the minimum pulse to pull-in the T90 relay, but may be adjusted to suit other types by altering $C_3$. Diode $D_{10}$ protects $Tr_5$ at switch-off.

In Figure 3, the initial voltage is 22 V per relay and the holding voltage 12 V, giving an initial power consumption of 1.85 W, falling to 960 mW long-term. The total power saving is just under a watt.

Running relays at a reduced holding voltage not only avoids the inelegance of consuming power for no good reason, but also speeds dropout time by reducing the magnetic energy stored. It could be argued that such a power saving is negligible. In a big Class-A amplifier it might be, but it makes sense in modest Class-B amplifiers idling for much of the time – which is of course almost all of them.

## Reference

1. Self, D. 'Precision Preamplifier '96', *Electronics World*, July/August, September 1996.

# 34 Cool audio power

*August 1999*

I have always thought that some kinds of information cry out to be presented graphically. A case in point is the ultimate destination of the power that an amplifier draws from its supply. Some goes usefully into the load, some is dissipated in the power semiconductors, and a little is lost in drivers, emitter resistors, and so on. All these quantities vary strongly with the percentage power output, and they overlap to give a confused and uninformative picture if plotted on a normal graph. The alternative method of power-partition diagrams, presented here, shows at once how much is power is taken and precisely where it goes; this is particularly important for more complex output stages, such as Class-G types. The idea is not wholly original; my inspiration came from what Victorian engineers called Sankey diagrams, a similar method of showing where the energy went in a large steam engine.

This chapter only deals with sinusoidal waveforms, for reasons explained at the start of the chapter. I feel bound to point that I was perfectly aware that sine waves and musical waveforms have little in common, but you have to start somewhere, and a waveform that allows you to easily check your simulator results against pure mathematics has some pretty strong advantages.

There are several important power relationships in designing an output stage. Both the average and peak power dissipated in the output devices must be considered when determining their type and number. The average power dissipated controls the heat-sink design.

In most amplifier types the power dissipation varies strongly with output signal amplitude as it goes from zero to maximum, so the information is best presented as a graph of dissipation against the fraction of the available rail-to-rail output swing – i.e., the output voltage fraction.

Consideration of average power allows the output devices to be made thermally safe; but it is also essential to consider the peak instantaneous dissipation in them. Audio waveforms have large low-frequency components, too slow for peak currents and powers to be allowed to exceed the DC limits on the data sheet.

For a resistive load the peak power is fixed and easily calculable. With a reactive load the peak power excursions are less easy to determine but highly important because they are increased by the changed voltage/phase relationships in the output device. Thus for a given load impedance modulus the peak power would need to be plotted against load phase angle as well as output fraction to give a complete picture.

Average power drawn from the rails is also a vital prerequisite for the power-supply design; since the rail voltage is substantially constant this can be easily converted into a current demand, which must be known when sizing reservoir capacitors, choosing rectifiers, and so on.

The voltage rating of these components is a much simpler business, requiring simply that they withstand the off-load voltages at the maximum mains voltage, which is usually taken as 10% above nominal. The only thing to decide is how big a safety margin is required.

Power drawn depends on signal-level and is again conveniently displayed with voltage-fraction as the X-axis.

## The mathematical approach

When dealing with power amplifier efficiency, most textbooks use a purely mathematical method as shown in Figure 1, which was produced with the aid of *Mathcad*. The calculation gives only the dissipation in the power devices.

Figure 1 gives the familiar information that maximum device dissipation occurs at 64% of maximum voltage, equivalent to 42% of maximum power. These specific numbers are a result of the sine waveform chosen and other waveforms give different values.

To make it mathematically tractable, the situation is highly idealised, assuming an exact 50% conduction period, no losses in emitter resistors or $V_{ce(sat)}$s, and so on. Solving the problem for Class AB, where the conduction period varies with signal amplitude, is considerably more complex due to the varying integration limits.

## Simulating dissipation

Alternatively, the power variations in real output stages can be simulated and the results plotted; the circuits simulated in this article are shown in Figure 2.

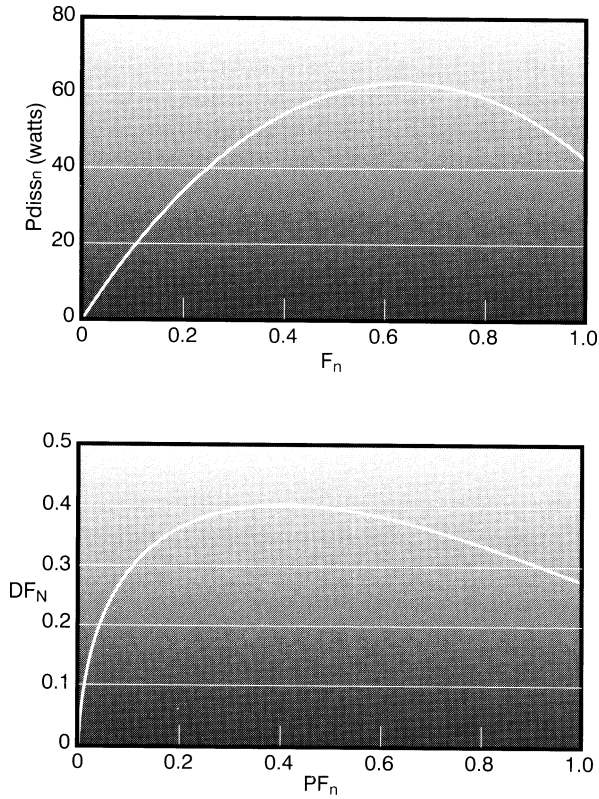



**Figure 1** *The standard mathematical derivation for Class-B. Maximum dissipation occurs at* 64% *voltage output, equivalent to* 42% *of maximum power output.*

For concision, and by analogy with logic outputs, I have called the upper transistor the source and the lower the sink. In simulation, losses and circuit imperfections are included, and the power dissipations in every part of the circuit, including power drawn from the supply rails, are made available by a single run.

It is an obvious choice – which I duly took – to use a sine waveform in the simulations. This allows a reality-check against the mathematical results. Reactive loads are easily handled, so long as it is appreciated that the simulation often has to be run for ten or more cycles to allow the conditions in the load to reach a steady state.

$F_n$ is the fraction of full output swing
$n = 0 \ldots 10$ $F_n = \frac{n}{10}$ $\omega = 1$ $t = 0$

Rail volts (± rails)
$V = 50$

Load resistance $R = 8$

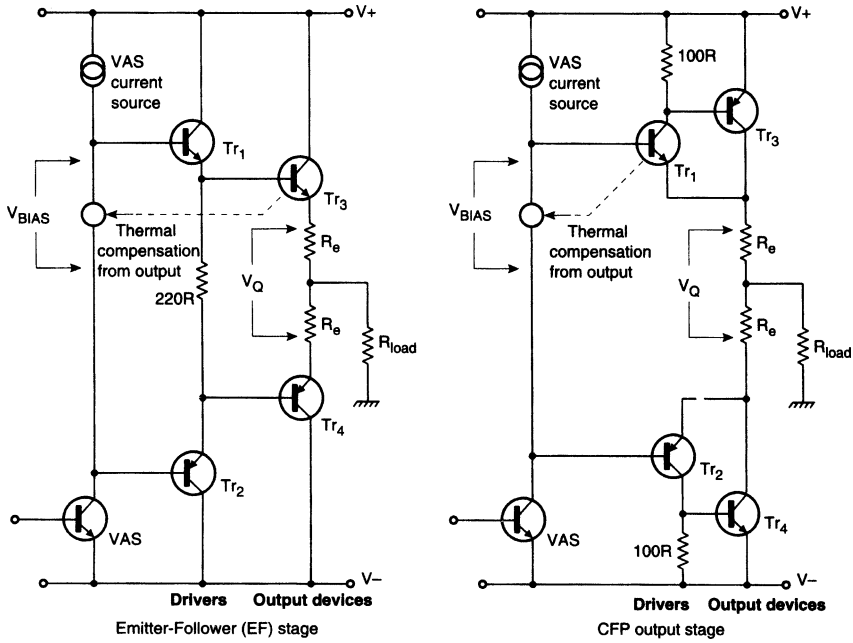



**Figure 2** *The standard emitter follower and complementary feedback pair output stages. In optimal Class-B the emitter follower version takes about* 150 mA *of quiescent current while the complementary feedback pair draws only* 10 mA.

$$Pout_n = \left[\frac{V}{\sqrt{2}}(F)_n\right]^2 \times \frac{1}{R} \quad \text{Output power}$$

Instantaneous power dissipation

$$P_n = \frac{V^2}{R}\left[F_n\left[\sin(\omega t) - F_n(\sin(\omega t))^2\right]\right]$$

$$Poutmax = \left[\frac{V}{\sqrt{2}}\right]^2 \times \frac{1}{R} \quad \text{Max. output power}$$

Integrate over one half cycle

$$Pdiss_n = \frac{V^2}{R}F_n\frac{1}{\pi}\int_0^x \sin(\omega t) - F_n(\sin(\omega t))^2\,dt$$

$$Poutmax = 156.25\,\text{W}$$

Note the $1/\pi$ due to integration from 0 to $\pi$

$$PF_n = \frac{Pout_n}{Poutmax} \quad \text{Output factor}$$

Since only one device conducts at once, dissipation for one is total dissipation

$$DF_n = \frac{Pdiss_n}{Poutmax} \quad \text{Dissipation factor}$$

All simulations were run with ±50 V rails and an 8 Ω resistive or reactive load. The output emitter resistors were 0.1 Ω. The drawback to this approach is that it is rather labour intensive. With my current simulation software, PSpice 6.0 for DOS, the steps are:

- Simulate the output stage over a whole cycle, for each input voltage fraction; 5% steps give enough points for a presentable curve; the •STEP command automates this.
- Display simulation results in the graphical post-processor. (In PSpice this is called PROBE) This assumes it can display computed quantities, e.g., $V_{ce} \times I_c$ to give instantaneous device-power. Peak and average results can be read from the same display as PROBE. There is a function called AVG, which – unsurprisingly – yields the running average over a cycle. This stage can be automated as a macro, which is just as well, since it has to be performed at least 20 times, once for each input fraction value.
- The awkward bit. The computed peak and averaged power dissipations at the end of the cycle are read out from the PROBE cursor and recorded by hand, for each value of input fraction. There seems to be no other way to extract the information.
- The data from the third step is typed into Mathcad, to produce the graphs shown in this article. Once the data has been entered, Mathcad can manipulate it in almost any way conceivable.

## Power-partition diagrams

The graph in Figure 1 gives only one quantity, the amplifier dissipation.

I suggest a more informative graph format that I call a power partition diagram, which shows how the input power divides between amplifier dissipation, useful power in the load, and losses in drivers, etc.

Power dissipations are plotted against the input voltage fraction; this is not quite the same as the output voltage fraction as these are real output stages with gain slightly less than one. The input fraction increases in steps of 0.05, stopping at 0.95 to avoid clipping. The X-axis may linear or logarithmic.

Figure 3 shows the power-partition diagram for a Class-B complementary feedback pair stage as in Figure 2, which has a low quiescent current. Line 1 plots the $P_{diss}$ in the sink (lower) device. Line 2 is source plus sink power. Line 3 is source plus sink plus load power.

The topmost line 4 is the total power drawn from the power supply, and so the narrow region between 3 and 4 is the power dissipated in the rest of the circuit – mainly the drivers and the output emitter resistors $R_e$. This power increases with output drive, but remains negligible compared with the other quantities examined.

The diagram shows immediately that the power drawn from the supply increases proportionally to the drive voltage fraction. This is partitioned

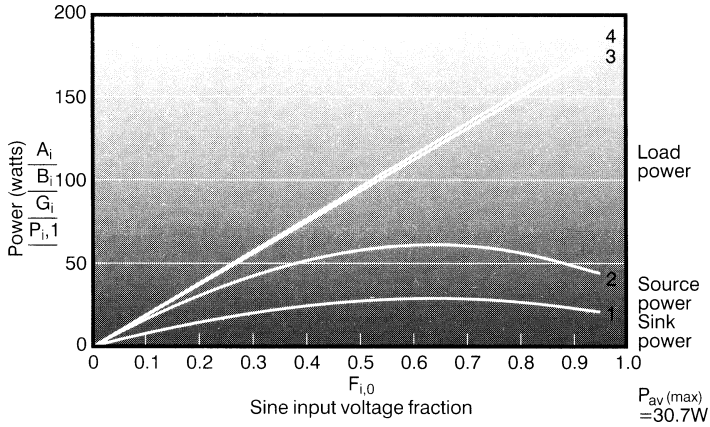



**Figure 3** *Power partition diagram for a Class-B complementary feedback pair driving an* 8 Ω *resistive load with a sinewave averaged over a cycle.*

between the load – represented by the curved region between lines 2 and 3 – and the output devices. Note how the peak in their power dissipation accommodates the curve of the load power as it increases with the square of the voltage fraction.

Figure 4 shows the same diagram for a Class-B emitter-follower output stage. The quiescent current of an emitter follower output stage is significant – here 150 mA – and pushes up the power dissipation around zero output, but at higher levels the curves are the same. There is no need for extra heatsinking over the complementary feedback pair case.

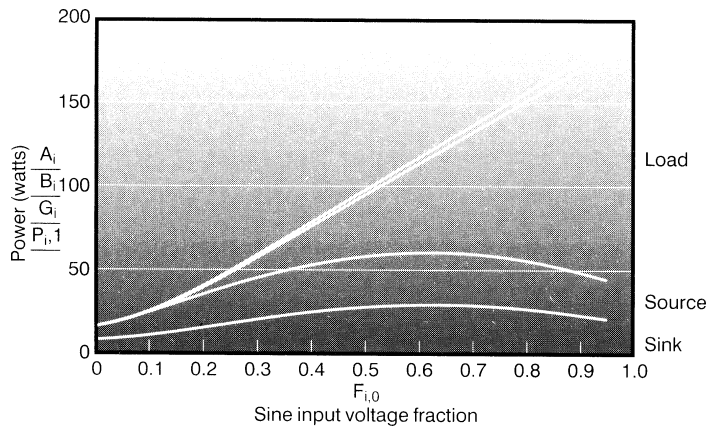



**Figure 4** *Class-B emitter follower power partition into* 8 Ω *resistive load. Sinewave drive. Significant quiescent dissipation at zero output.*

## Effects of increased bias

Figure 5 shows Class-AB, with bias increased so that Class-A operation and linearity is maintained up to 5 W r.m.s. output.

The quiescent current has increased to 370 mA, so quiescent power dissipation is significantly higher for output fractions below 0.1. Device dissipation is still greatest at a drive fraction of around 0.6, so once again no extra cooling is required to deal with the increased quiescent dissipation.

A push-pull Class-A amplifier draws a large standing current, and the picture looks totally different; see Figure 6. The power drawn from the supply is constant, but as output increases dissipation transfers from the output devices to the load, so minimum amplifier heating is at maximum output.

The significant point is that amplifier dissipation is only meaningfully reduced at a voltage fraction of 0.5 or more, i.e. only 6 dB from clipping. Compared with Class-B, an enormous amount of energy is wasted internally.

Single-ended or constant-current versions of Class-A have even lower efficiency, worse linearity, and no corresponding advantages.

### *Class G*

Hitachi introduced the Class-G concept in 1976 with the aim of reducing amplifier power dissipation by exploiting the high peak-mean ratio of music.[1] I have recently explained its operation in Ref. 2.

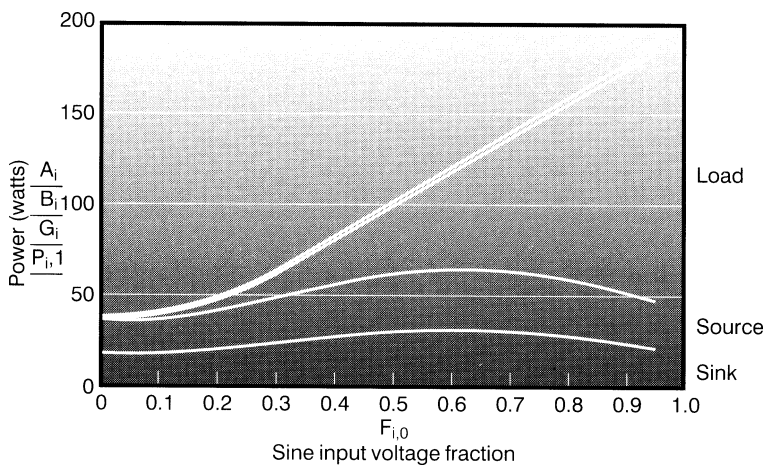



**Figure 5** *Class-AB sinewave drive. If the range of Class-A operation is extended, the area below level = 0.1 advances upwards and to the right.*

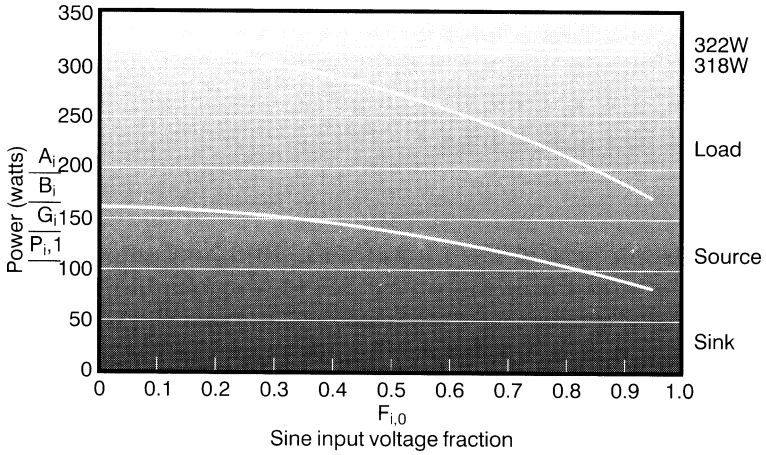



**Figure 6** *Class-A push-pull sinewave drive. Almost all the power drawn is dissipated in the amplifier, except at the largest outputs.*

At low outputs, power is drawn from a pair of low-voltage rails; for the relatively infrequent excursions into high power, higher rails are drawn from. Here the lower rails are ±15V, 30% of the higher ±50V rails, so I call this Class G (30%).

This gives a discontinuous power-partition diagram, as in Figure 7. Line 1 is the dissipation in the low-voltage inner source device, which is kept low by the small voltages across it. Line 2 adds the dissipation in the high-voltage outer source; this is zero below the rail-switching threshold.

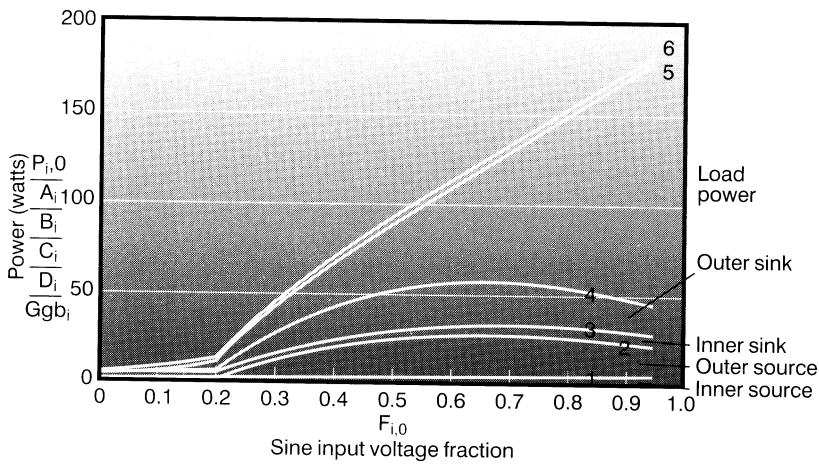



**Figure 7** *Class-G with lower supply rails set at* 30% *of upper rails, with sinewave drive. Compare Figure 4; amplifier dissipation at low levels is much reduced.*

Above this are added the identical – due to symmetry – dissipations in the inner and outer sink devices, as Lines 3 and 4. Line 5 adds the power in the load, and Line 6 is the total power drawn, as before.

Power consumption and amplifier dissipation at low outputs are much reduced; above the threshold these quantities are only slightly less than for Class-B. Class-G does not show its power-saving abilities well under sinewave drive.

## Class-B and reactive loads

The simulation method outlined above is also suitable for reactive loads. It is however necessary to run the simulation not just for one cycle, but sometimes for as many as twenty. This is to ensure that steady-state conditions have been reached.

The diagrams referred to below are for steady-state 200 Hz sinewave drive; the frequency must be defined so the load impedance can be set by suitable component values, but otherwise makes no difference.

Figure 8 shows what happens in Class-B emitter follower when driving a 45° capacitive-reactive load with a modulus of 8 Ω. Comparing it with Figure 4, the power drawn from the supply is essentially unchanged, and is still proportional to output voltage fraction.

The larger areas at the bottom show that more power is being dissipated in the output devices and correspondingly less in the load, because the phase shift causes the voltage across and the current through the output

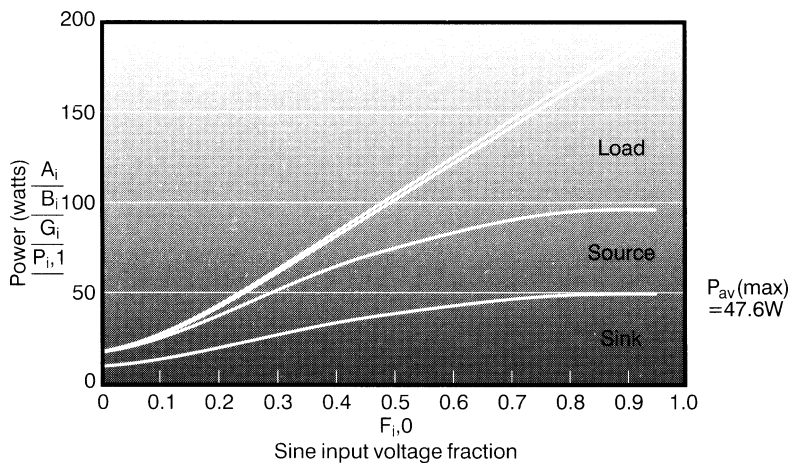



**Figure 8** *Class-B emitter follower reactive* 45°, *sinewave drive. The load is* 11.3 Ω *in parallel with* 71 μF. *Impedance modulus is* 8 Ω *at* 200 Hz. *Amplifier dissipation is increased, power delivered to the load decreased.*

devices to overlap more. The amplifier must dispose of 95 W of heat worst-case, rather than 60 W.

Average device dissipation no longer peaks, but increases monotonically up to maximum output. 45° phase angles are common when loudspeakers are driven. It is generally accepted that an amplifier should be able to provide full voltage swing into such a load.

When the load is purely reactive, with a phase angle of 90°, it can dissipate no power and so all that delivered to it is re-absorbed and dissipated in the amplifier. Figure 9 shows that the worst-case device dissipation is much greater at 185 W, absorbing all the power drawn from the supply, and therefore necessarily increasing monotonically with output level; there is no maximum at medium levels.

This is a very severe test for a power amplifier. It is also unrealistic, as no assemblage of moving-coil speaker elements can ever present a purely reactive impedance; 60° loads are normally the most reactive catered for. Table 1 shows the worst-case cycle-averaged dissipation for various load angles, showing how the position of maximum dissipation moves towards full output as the angle increases.

This is best displayed in 3D, as Figure 10, which plots power vertically; the slight hump at the front – non-reactive load – disappearing as the load becomes more reactive.

The dissipation hump is of little practical significance. An audio amplifier will almost certainly be required to drive 45° loads, and these cause higher power dissipations than resistive loads driven at any level.

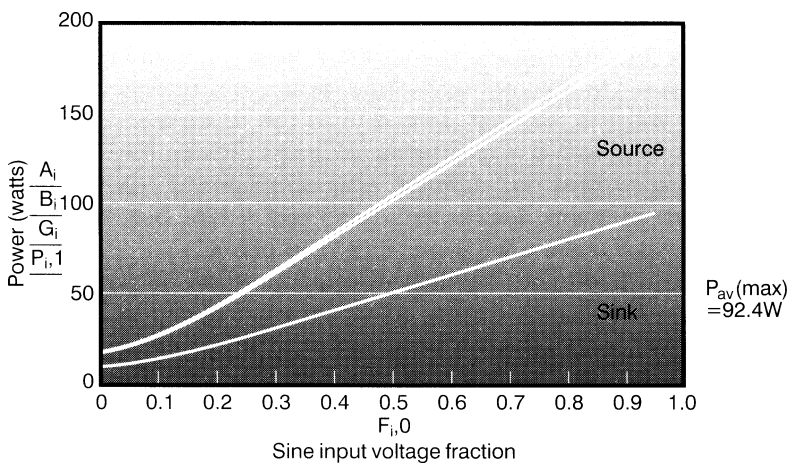



**Figure 9** *Class-B emitter follower reactive* 90°*, sinewave drive. Load is a* 99.5 μF *capacitor. Impedance modulus still* 8 Ω *at* 200 Hz*. All the supply power is now being absorbed by the amplifier, and none by the load.*

**Table 1**

| *Angle* (°) | $P_{diss(max)}$ (W) | *Voltage fraction* |
|---|---|---|
| 0 | 60 | 0.64 |
| 10 | 63 | 0.65 |
| 20 | 67 | 0.70 |
| 30 | 70 | 0.75 |
| 45 | 95 | 0.95 |
| 60 | 115 | 1.00 |
| 90 | 185 | 1.00 |

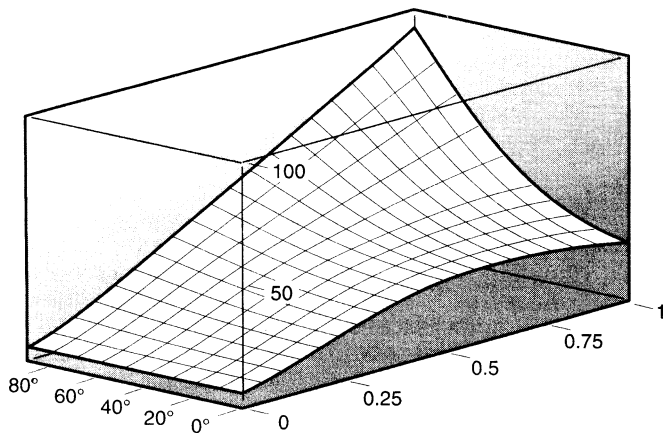



**Figure 10** *The average power (vertical axis) against load angle (left-hand horizontal axis) and output fraction (right-hand horizontal axis).*

Figure 11 shows the same plot for peak power, which increases monotonically with both output fraction and load angle. Figure 12 summarises all this data for design purposes. It shows worst-case peak and average power in one output device against load reactance.

Peak powers are taken at 0.95 of full output, average power at whatever output fraction gives maximum dissipation. Therefore to design an amplifier to cope with 45° loads, note that average power is increased by 1.4 times, and peak power by 2.7 times, over the resistive case. This can mean that it is necessary to increase the number of output devices simply to cope with the much enhanced peak power.

Considering simple reactive loads like those listed in the panel 'Reactive load observations' gives an essential insight into the extra stresses they impose on semiconductors but is still some way removed from real signals

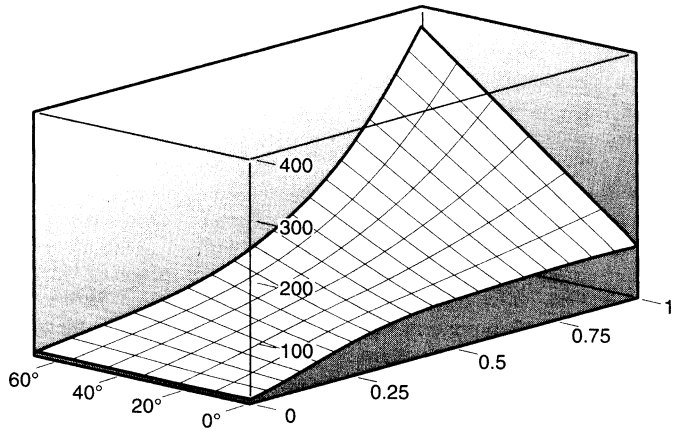



**Figure 11** *Peak power plotted as in Figure 10. The vertical scale must accommodate much higher power levels than Figure 10.*

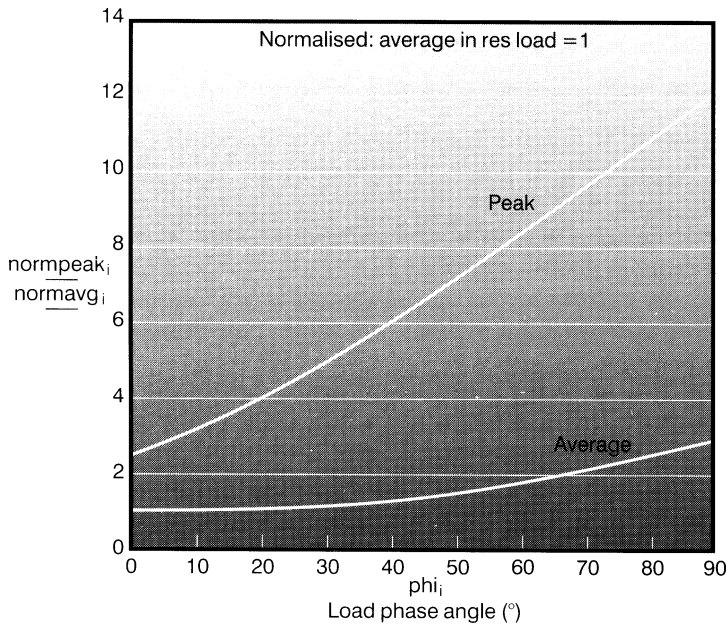



**Figure 12** *Peak power increases faster than worst-case average power as the load becomes more reactive and its phase angle increases. Class-B emitter follower as before.*

and real loudspeaker loads, where the impedance modulus varies along with the phase, due to electromechanical resonances or crossover dips.

I looked at single and two-unit loudspeaker models in Ref. 3 where the maximum phase angle found was 40°. In brief, the results were:

- Amplifier power consumption and average supply current drawn vary with frequency due to impedance modulus changes.
- The peak device current increased by a maximum of 1.3 times at the modulus minima.
- The average current in the output devices increased by a maximum of 1.3 times.
- Peak device power increased by a maximum factor of 2, mostly due to phase shift rather than impedance dips.
- Average device dissipation increased by a maximum of 1.4 times.

## Reactive load observations

The following conclusions apply to reactive loads.

- Amplifier power consumption and average supply current drawn do not vary with load phase angle if the impedance modulus remains constant.
- Peak device current is not altered so long as the impedance modulus remains constant.
- Average current in the output devices is not altered so long as the impedance modulus remains constant. This follows from the first observation.
- Peak device power increases rapidly, as the load becomes more reactive. A 45° load increases power peaks by 2.7 times, and a 60° load by 3.4 times. See Figure 12.
- Average device dissipation also increases, but more slowly, as the load angle increases. A 45° load increases average dissipation by 1.4 times and a 60° load by 1.8 times, Figure 12.

These numbers come from two specific models that attempted to represent 'average' speakers. Worse conditions could easily have been found.

Ultimately a comprehensive survey of the loudspeakers on the market would be required, but this would be very time-consuming. In Ref. 4, which gives an excellent account of real speaker loading, 21 models were tested and the worst angle found was 67°. Eliminating the two most extreme cases reduced this to 60°.

The most severe effect of reactive loads is the increase in peak power, followed by the increase in average power.

Both are a strong function of load phase, and so the specification of the maximum angle to be driven has a big effect on the devices required, heat-sink design, and hence on amplifier cost.

It is likely that a failure to appreciate just how quickly peak power increases with load angle is the root cause of many amplifier failures.

## References

1. Feldman, 'Class-G high efficiency hi-fi amplifier', *Radio-Electronics*, August 1976, p 47.
2. Self, D. 'Classification of power amplifiers', *Electronics World*, March 1999, p 90.
3. Self, D. 'Speaker impedance matters', *Electronics World*, November 1997, p 920.
4. Benjamin, E. 'Audio power amplifiers for loudspeaker loads', *Journal Audio Engineering Society*, Vol. 42(9), September 1994, p 670.

# 35 Audio power analysis

*December 1999*

This article extended the idea of power-partition diagrams to include real waveforms of real music, and therefore required rather a lot of tedious measurement to acquire the probability density data. One of the surprises was that very different musical genres had very similar statistics. It was however reassuring to discover that the demands made on an amplifier by a sinewave were greater than those for music, but not wildly out of line. In other words, sinewave testing builds in a safety margin – if it works on sinewaves it will certainly survive music. Sinewave testing has received a lot of unkind criticism over recent years, but it continues unabated for the simple reason that it works.

Writing this put me off designing Class-A amplifiers for a long time. A branch of technology with a practical energy efficiency of less than 1% needs a good justification for its existence, and the distortion from a Blameless Class-B amplifier is so low that the extra linearity of Class-A amplifiers is no longer the forceful argument it was.

My last chapter showed how the power consumed by amplifiers of various classes was partitioned between internal dissipation and the power delivered to the load.[1] This was determined for the usual sinewave case.

The snag with this approach is that a sinewave does not remotely resemble real speech or music in its characteristics. In many ways it is almost as far from it as you could get.

In particular, it is well-known that music has a large peak-to-mean ratio, or PMR, though the actual value of this ratio in decibels is a vague quantity. Signal statistics for music appear to be in surprisingly short supply.

Very roughly, general-purpose rock music has a PMR of 10 dB to 30 dB, while classical orchestral material – which makes very little use of fuzz boxes and the like – is 20 to 30 dB. The muzak you endure in lifts is limited in

PMR to 3 to 10 dB, while compressed bass material in live PA systems is similar.

It is clear that the power dissipation in PA bass amplifiers is going to be radically different from that in hi-fi amplifiers reproducing orchestral material at the same peak level. The PMR of a sinewave if 4.0 dB, so results from this are only relevant to lifts.

Recognising that music actually has a peak-to-mean ratio is a start, but it is actually not much help as it reduces the statistics of signal levels to a single number. This does not give enough information for the estimation of power dissipation with real signals.

To calculate the actual power dissipations, two things are needed; a plot of the instantaneous power dissipations against level, and a description of how much time the signal spends at each level. The latter is formally called the 'probability density function', or PDF, of the signal; more on this later.

The instantaneous power partition diagram, or IPPD, is obtained by running the output stage simulation with a sawtooth input and no per-cycle averaging. Instantaneous power dissipation can therefore be read out for any input voltage fraction simply by running the cursor up the sawtooth.

Figure 1 is the instantaneous power partition diagram for the Class-B complementary-feedback pair case, where the quiescent current is very small. This looks very much like the averaged-sinewave power partition diagram in reference,[1] but with the device dissipation maximum at 50% voltage rather than 64% for the sinewave case.

The instantaneous powers are much higher, as they are not averaged over a cycle. There is only one device-power area at the bottom as only

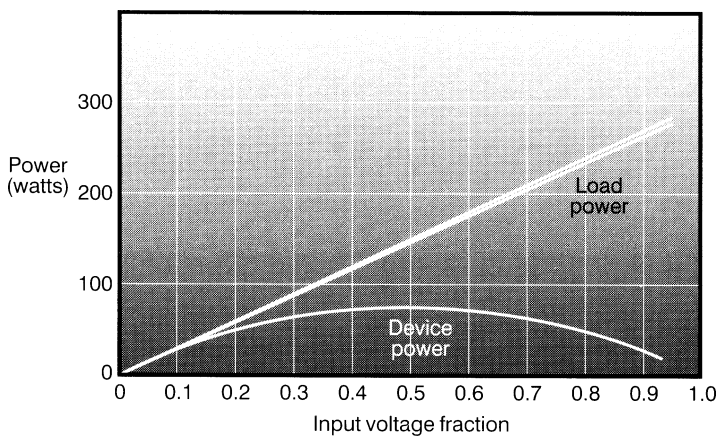



**Figure 1** *Instantaneous power partition diagram for Class-B complementary-feedback pair. Power in the output devices peaks when output is at half the rail voltage.*

one device conducts at a time. Output device dissipation at the moment when the signal is halfway between rail and ground, input fraction 50% – is 76 W, and the power in the load is 75 W. This total to 151 W, on the lower of the two straight lines, while the power drawn from the supply is shown as 153 W by the upper straight line. The 2 W difference represents losses in the driver transistors and the output emitter resistors.

All the IPPDs for various output stages look very similar in shape to the averaged-sine PPDs in Ref. 1, but the peak values on the Y-axis are higher. The IPPD can be combined with any PDF to give a much more realistic picture of how power dissipation changes as the level of a given type of signal is altered.

## The probability density function

The most difficult part of the process above is obtaining the probability density function. For repetitive waveforms the PDF can be calculated,[2] but music and speech need a statistical approach. It is often assumed that musical levels have a Gaussian (normal) probability distribution, as the sum of many random variables.

Positive statements on this are however hard to find. Benjamin[3] says, 'music can be represented accurately as a Gaussian distribution' while Raad[4] states, 'music and mixed sounds typically have Gaussian PDFs'. It appears likely this assumption is true for multi-part music which can be regarded as a summation of many random processes; whatever the PDF of each component, the result is always Gaussian as indicated by the Central Limit Theorem.

If the distribution is Gaussian, its mean is clearly zero, as there is no DC component, which leaves the variance – i.e. width of the bell-curve – as the only parameter left to determine. The Gaussian distribution tails off to infinity, implying that enormous levels can occur, though very rarely.

In reality the headroom is fixed. I have dealt with this by setting variance so the maximum value, 0 dB. occurs 1% of the time. This is realistic as music very often requires judicious limiting of occasional peaks to optimise the dynamic range.

The PDF presents some conceptual difficulties, as it shows a density rather than a probability. If a signal level ranges between 0 and 100%, then clearly it might be expected to spend some of its time around 50%.

However, the probability that it will be at exactly 50.000% is zero, because a single level value has zero extent. Hence the PDF at $x$ is the probability that the signal variable is in the interval $(x, x + dx)$, where $dx$ is the usual calculus infinitesimal.

## The cumulative distribution function

If the probability that the instantaneous voltage will be above – not at – a given level is plotted against that level, a cumulative distribution function, or CDF, results. This is important as it is easier to measure than the PDF.

If the variable is $x$, then the PDF is often called $P(x)$ and the CDF called $F(x)$. These are related by:

$$P(x) = \frac{\mathrm{d}}{\mathrm{d}x} F(x)$$

or,

$$F(x) = \int_0^x P(a)\,\mathrm{d}a$$

where $a$ is a dummy variable needed to perform the integration. The integration starts at zero in this case because signal levels below zero do not occur.

Generating a CDF by integrating a given PDF is straightforward, but going the other way – determining the PDF from the CDF – can be troublesome as the differentiation accentuates noise on the data.

## Some probability density functions

Figure 2 shows the calculated PDF of a sinewave. As with every PDF, the area under the curve is one, because the signal must be at some level all of the time.

However, the function blows up – i.e. heads off to infinity – at each end because the peaks of the wave are 'flat', and so the signal dwells there for infinitely longer than on the slopes where things are changing. These

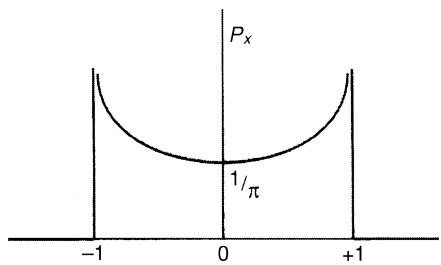



**Figure 2** *Probability density function, or PDF, of a sinewave. Peaks at each end go towards infinity.*

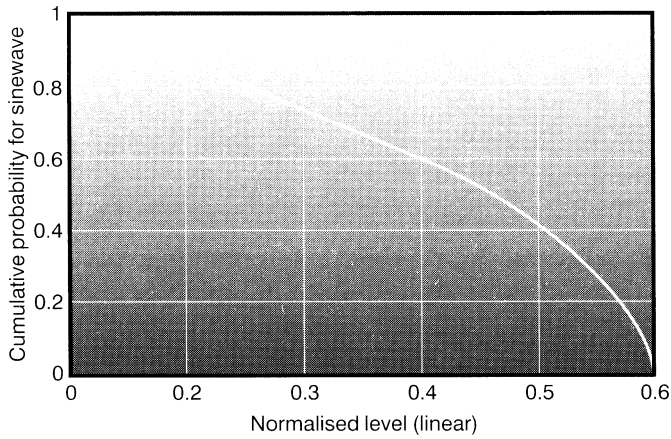



**Figure 3** *Cumulative distribution function, CDF, of a sinewave. Drawn with measured data from the circuit of Figure 4 as a reality check.*

'flat' bits are infinitely small in time extent though, and so the area under the curve is still unity. This shows you why PDFs are not always the easiest things to handle.

The CDF for a sinewave is shown in Figure 3; the probability of exceeding the level on the axis falls slowly at first, but then accelerates to zero as the rounded peaks are reached.

## Measuring probability density functions

But is all music Gaussian? I was not satisfied that this had been conclusively established from just two brief references.

I decided it was essential to make some attempt to determine musical PDFs. In essence this is simple. The first thing to decide is the length of time over which to examine the signal. For most contemporary music the obvious answer is 'one track', a complete composition lasting typically between three and eight minutes.

Very simple circuitry can be used to determine a CDF, and hence the PDF, though the process is protracted. A variable-threshold comparator is driven by the signal to be measured, and its output applied to a long-period averaging time-constant. Figure 4.

A comparator, $IC_{1a}$, rather than an op-amp, is used to avoid inaccuracies due to slew-rate limiting. Reference $IC_2$ is an inexpensive 2.56 V bandgap type, while $VR_1$ sets the comparator threshold. When the signal level is below this threshold, the comparator open-collector output is off, and the voltage seen by the averaging network is zero.

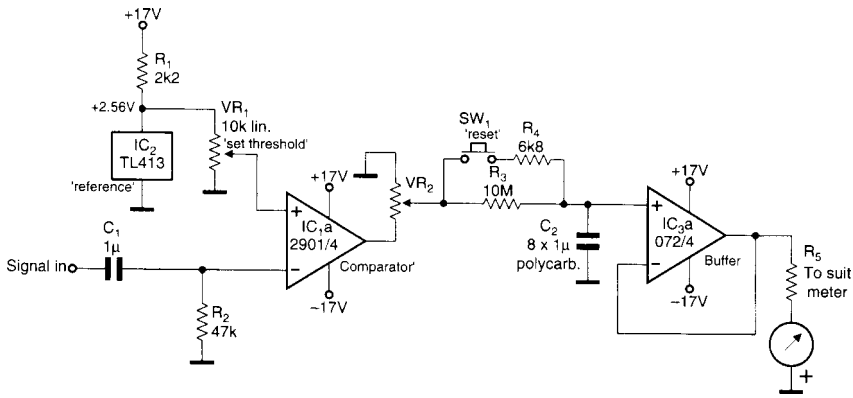



**Figure 4** *Simple circuit for measuring the CDF of an audio signal.*

When the signal exceeds threshold, the comparator output is pulled low, so this point carries an irregular rectangular waveform while signal is applied. The average value of this is derived by $R_3$ and $C_2$, buffered by $IC_{3a}$, and drives a moving-coil meter through a suitable resistance $R_5$.

Switch $SW_1$ and $R_4$ enable a quick reset when no signal is present. A moving-coil meter allows much easier reading of a changing signal, though not to any great accuracy.

Potentiometer $VR_2$ sets the scale so that the meter deflects to full scale for a 100% reading. This is done with no input, so it is essential to check that the circuit offsets have put the comparator in the right state – i.e. output low; if not the inverting input will need to be pulled fractionally negative by a high-value biasing network.

The circuit only measures one polarity of the waveform, in this case the positive half, so signal symmetry is assumed. This is safe unless you plan to do a lot of work with solo instruments or single a cappella voices; the human vocal waveform is notably asymmetrical.

This minimal system is simple, but it only yields one data point at a time. Set the threshold level to say 50%, play the track – I'd pick a short one – and as it finishes the reading on the meter shows the percentage of time the signal exceeded the preset level.

Since twenty data points are required for a good graph, this gets pretty tedious. The four comparators of $IC_1$ could give four points, if the time-constant section was also quadrupled, and some means of freezing the output voltages provided.

The CDF thus obtained for Alannah Myles' 'Black Velvet' is Figure 5, and the PDF derived from it is Figure 6. It comes complete with some

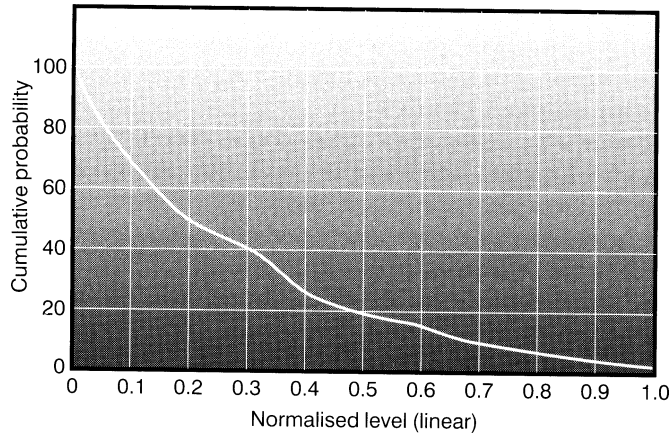



**Figure 5** *The cumulative distribution function obtained from the PDF of Alannah Myles performing 'Black Velvet' by Figure 4.*

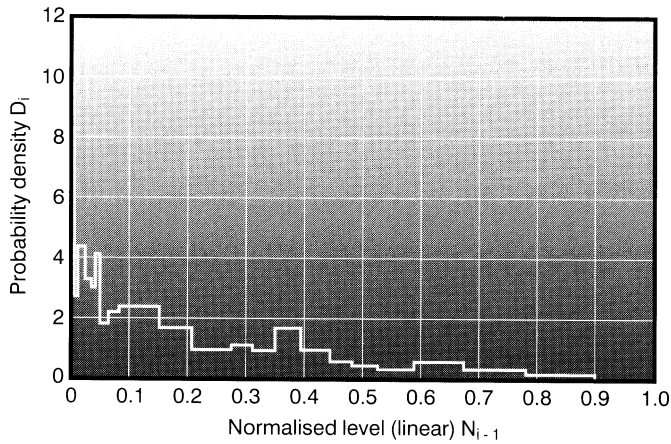



**Figure 6** *Probability density function derived from Alannah Myles performing 'Black Velvet'.*

rather implausible ups and downs produced by differentiating data that is accurate to ±1% at best.

I measured several rock tracks, and also short classical works by Albinoni and Bach. The results are surprisingly similar; see the composite CDF in Figure 7. This is good news because we can use a single PDF to evaluate amplifiers faced with varying musical styles. However, I decided the method needed a reality check, by deriving the PDF in a completely different way.

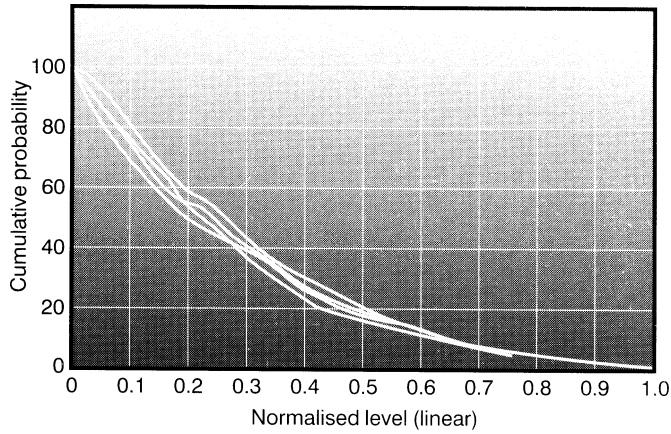



**Figure 7** *The CDFs of 3 rock and 1 classical tracks, showing only small differences.*

## Probability density functions via DSP

A digital processor offers the possibility of determining as many data points as you want on one playing of the music specimen. In this case a very simple 56001 program sorts the audio samples into 65 amplitude bins.

The result for 30 s of disco music is Figure 8, which is somewhere between triangular and Gaussian, if the latter has appropriate variance. The important point is that the difference between them is very small, and either

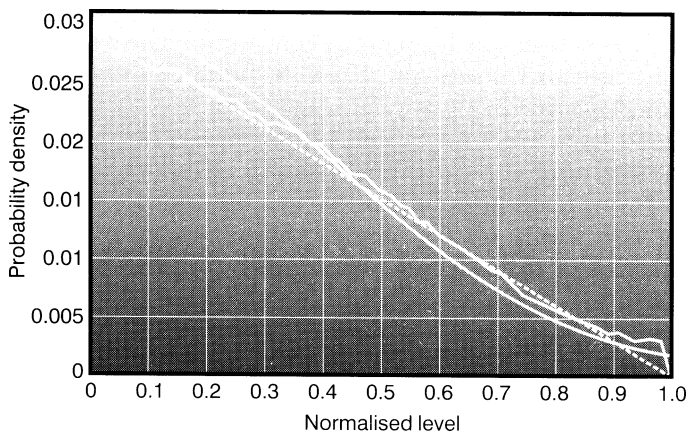



**Figure 8** *The PDF of disco music, sorted into 65 amplitude bins by DSP. Also shown are a Gaussian distribution (smooth curve) and a triangular distribution (dotted line).*

can be used. The triangular PDF simplifies the mathematics, but if like me you use *Mathcad* to do the work, it is easy to plug in whatever distribution seems appropriate.

## Deriving actual power

Having found the PDF, it is combined with the power partition diagram. In this case the IPPD is divided into twenty steps of voltage fraction, and each one multiplied by the probability the signal is in that region.

The summation of these products yields a single number – the average power dissipation in watts for a real signal that just reaches clipping for 1% of the time. An obvious extension of the idea is to plot the average power derived as above, against signal level on the X-axis. This gives an immediate insight into how amplifer power varies as the general signal level is reduced, as by turning down the volume control.

Figure 9 shows how level changes affect the PDF. Line 1 is maximum volume, just reaching full volume at the right. Line 2 is half volume, −6 dB, and so hits the X-axis at 0.5; it is above Line 1 to the left as the probability of lower levels must be higher to maintain unity area under it.

This process continues as volume is reduced, until at zero volume the zero-level probability is 1 and all other levels have zero probability. Having generated twenty PDF functions, the powers that result for each one are plotted with the volume setting – not the output fraction – as the X-axis. The results for some common amplifer classes are as follows.

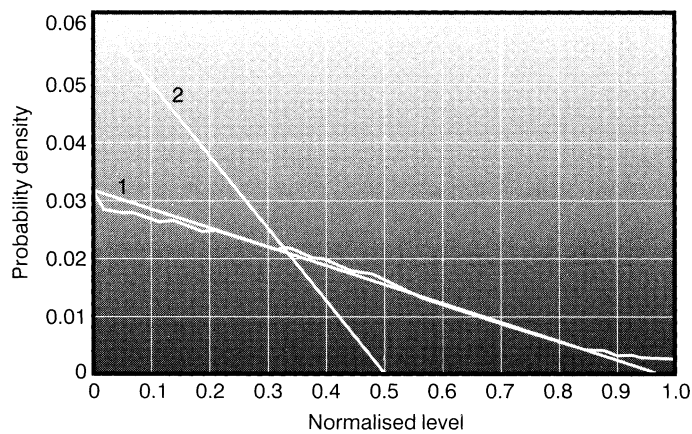



**Figure 9** *The triangular PDF, and how level changes affect it. Line 1 is full volume, and Line 2 half volume.*

### Class-B

The instantaneous power plot for Class-B complementary feedback pair combined with a triangular PDF of Figure 10 illustrates how the load and device power varies with volume setting. A signal with triangular PDF spends most of its time at low values, below 0.5 output fraction, and so there is no longer a dissipation maximum around half output. Device dissipation at bottom increases monotonically with volume. Load power increases with a square-law, which is a reassuring check on all these calculations.

Figure 11 is Figure 10 replotted with a logarithmic X-axis, which is more applicable to human hearing. Domestic amplifiers are rarely operated on the edge of clipping; a realistic operating point is more like −15 or −10 dB. The plot reveals that here the efficiency is low, with much more power dissipated in the devices than reaches the load.

### Class-AB

A decibel plot for Class-AB, biased so Class-A operation is maintained up to 5 W r.m.s. output is shown in Figure 12. Quiescent current is now 370 mA, so there is greater quiescent dissipation at zero volume. There is also substantial conduction overlap, and so sink and source would be different if the plot only considered voltage excursions in one direction away from 0 V. When positive and negative half-cycles are averaged, symmetry is achieved. The total device dissipation is unchanged but the boundary between the source and sink areas is half way, as in Figure 12.

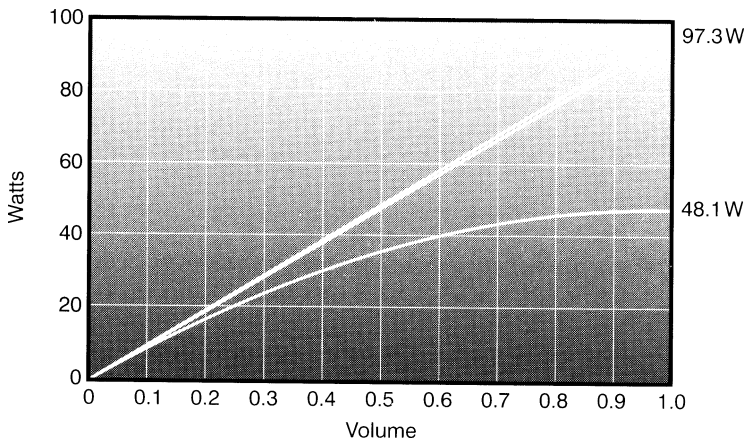



**Figure 10** *Class-B complementary feedback pair power partition versus level. The Class-B IPPD has been combined with the triangular PDF. Device dissipation (lower area) now increases monotonically with volume.*

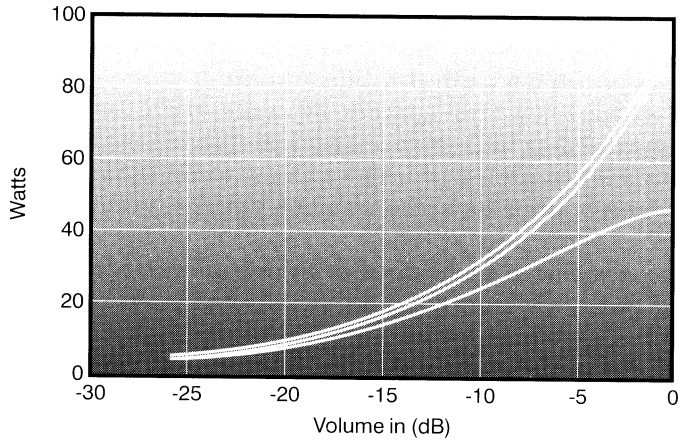



**Figure 11** *Class-B complementary feedback pair plotted with volume on a more useful logarithmic (decibel) X-axis. The shape looks quite different from Figure 10.*

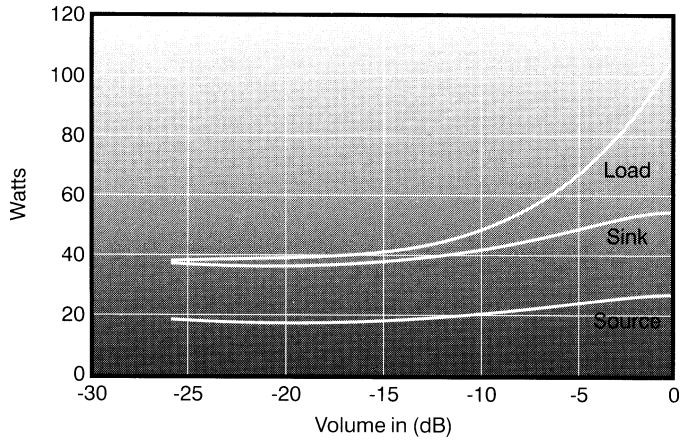



**Figure 12** *Class-AB Power Partition Diagram, stage biased to give Class-A up to* 5 W. *Averaged over whole cycle.*

## *Class-A push-pull*

I have stuck with the same ±50 V rails for ease of comparison, and this yields a very powerful Class-A amplifier. The power drawn from the load is constant, and as output increases dissipation transfers from the output devices to the load, giving minimum amplifier heating at maximum output.

The result for sinewave drive is bad enough,[1] but Figure 13 reveals that with real signals, almost all the energy supplied is wasted internally – even

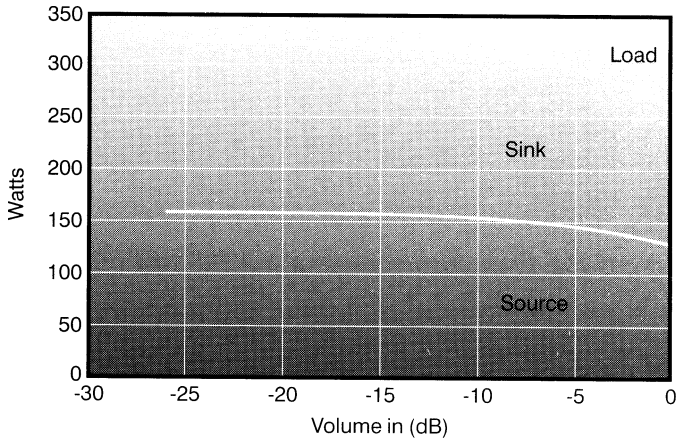



**Figure 13** *Class-A push-pull, for* 150 W *output. The internal dissipation completely dominates – even at maximum volume.*

at maximum volume. Class-A has always been stigmatised as inefficient; this shows that under realistic conditions it is hopelessly inefficient, so much so that it grates on my sense of engineering aesthetics. At typical listening volumes of −15 dB the efficiency barely reaches 1%.

## *Class-G*

This class of amplifiers was introduced by Hitachi in 1976 to reduce amplifier power dissipation by exploiting the high peak-mean ratio of music.[5] Class-G made little headway in the hi-fi market as the power saving does not outweigh the increased circuit complexity, but the rise of five-channel home theatre applications has caused a revival of interest in improved amplifier efficiency.

I recently explained Class-G in Ref. 6. At low outputs, power is drawn from low-voltage rails; for the relatively infrequent excursions into high power, higher rails are switched in.

In Figure 14 the lower rails are ±15 V, 30% of the higher ±50 V rails; I call this Class-G-30%. The lower area is the power in the inner devices – i.e. those in all the time. The larger area just above is that in the outer devices, i.e. those only activated when running from the higher rails. This is zero below the rail-switching threshold at a volume of 0.2.

Total device dissipation is reduced from 48 W in Class-B to 40 W, which is not a good return for twice as many power transistors. This is because the lower rail voltage is poorly chosen for signals with a triangular PDF.

If the low rails are increased to ±30 V this become Class-G-60% as in Figure 15. Here the low-dissipation region now extends up to a voltage

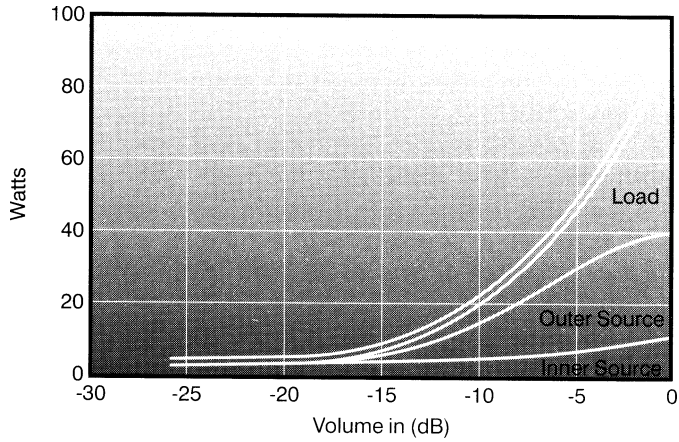



**Figure 14** *Class-G-30%. Low rail voltage is 30% of the high rail. Rail-switching occurs at about −15 dB relative to maximum output.*

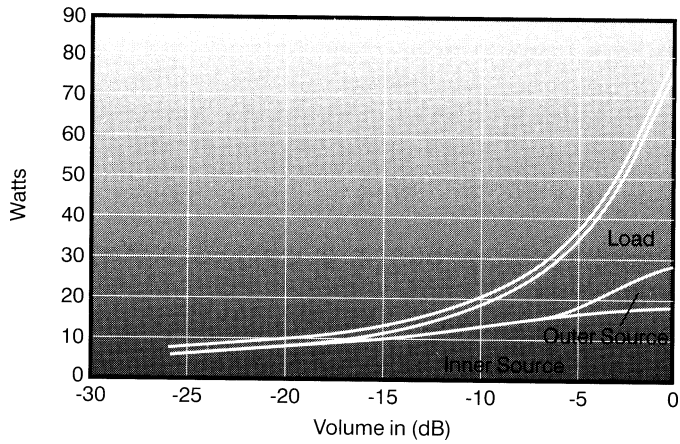



**Figure 15** *Class-G-60%. The low rail voltage is now 60% of the high rail. This reduces both dissipation and power consumed, compared with Figure 14.*

fraction of 0.5, but inner device dissipation is higher due to the increased lower rail voltages.

The overall result is that total device power is reduced from 48 W in Class-B to 34 W, which is a definite improvement. I am not suggesting that 60% is the optimum lower-rail voltage. The efficiency of Class-G amplifiers depends very much on signal statistics.

## Reactive loads

The disadvantage of using instantaneous power is that it ignores signal and circuit history, and so cannot give meaningful information with reactive loads. The peak dissipations that these give rise to with real signals are difficult to simulate; it would be necessary to drive the circuit with stored music signals for many cycles; and that would only cover a few seconds of a CD or concert. The anomalous speaker currents examined in Ref. 7 show how significant history effects can be with some waveforms.

## In summary

Tables 1 and 2 summarise how a triangular-PDF signal – rather than a sinewave – reduces average power dissipation, and the power drawn from the supply.

These economies are significant; the power amplifier market is highly competitive, and it is essential to exploit the cost savings in heat-sinks and ower-supply components made possible by designing for real signals rather than sinewaves.

In particular, Class-G shows valuable economies in device dissipation and ower-supply capacity, though to reduce dissipation, the lower supply voltage must be carefully chosen. This approach is unlikely to reduce the

**Table 1** *Device dissipation, worst-case volume*

| | *Sinewave* (W) | *PDF* (W) | *Factor* |
|---|---|---|---|
| Class-B CFP | 64 | 48 | 0.75 |
| Class-AB | 64 | 55 | 0.78 |
| Class-A, push-pull | 324 | 324 | – |
| Class-G-30% | 43 | 40 | 0.93 |
| Class-G-60% | 56 | 34 | 0.61 |

**Table 2** *Power drawn, worst-case. Always maximum output*

| | *Sinewave* (W) | *PDF* (W) | *Factor* |
|---|---|---|---|
| Class-B CFP | 186 | 97 | 0.52 |
| Class-AB | 188 | 105 | 0.58 |
| Class-A, push-pull | 324 | 324 | – |
| Class-G-30% | 177 | 93 | 0.52 |
| Class-G-60% | 169 | 81 | 0.48 |

number of power devices required as real signals give no corresponding reduction in peak device power or peak device current.

## References

1. Self, D. 'Cool audio power', *Electronics World*, August 1999, p 657.
2. Carlson, A. 'Communication systems', McGraw-Hill, Third Edn 1986, p 144.
3. Benjamin, E. 'Audio power amplifiers for loudspeaker loads', *Journal Audio Engineering Society*, Vol. 42, September 1994, p 670.
4. Raad, 'Average efficiency of class-G amplifiers', *IEEE Trans on Consumer Electronics*, Vol CE-22(2), May 1986.
5. Feldman, 'Class-G high efficiency hi-fi amplifier', *Radio-Electronics*, August 1976, p 47.
6. Self, D. 'Class Distinction: amplifier classes', *Electronics World*, March 1999, p 190.
7. Self, D. 'Loudspeaker undercurrents', *Electronics World*, February 1988, p 98.

# Index

## PREAMPLIFIERS & MIXERS


Absolute phase, 34, 54
Active rectifiers, 122
Active volume control, 14, 25, 26, 52, 76
Analog gates, 139, 144
  control voltage feedthrough, 154
  in current mode, 151
  distortion of, 154
  paralleling, 151
  series-shunt, 153
  series switching, 146
  voltage mode, 146
Attack time, 123–5
Attenuator, 133
Auxiliary sends, 134

Balance control, 23, 32, 69
Balanced line inputs, 104
Balanced line outputs, 99
Balanced mix bus, 137
Bass frequency control, 73
Baxandall, Peter, 7, 26, 33, 128
Bootstrapping:
  collector loads, 22
Buffer stage, 31, 34, 52
Bus residual, 137

Channel muting, 139
Clipping detection, 8
Close-tolerance capacitors, 30, 60
Common-impedance coupling, 95
Common-mode rejection (CMRR), 96, 104, 133
Compressor, 117–26
Crosstalk, 35, 39, 67, 83, 112, 137

DC blocking, 42, 82, 157
Decay time, 123–5
Decoupling, 34
Differential amplifier, 105
Differential pair input, 42
Distortion, 146, 158

Effect returns, 135
Electrostatic coupling, 93
End-stop resistors, 73
EQ, 134

FETs, 118
  distortion cancellation, 120
  switching, 139
Foldback, 134, 135
Frequency response, 3

Galvanic isolation, 102, 143
Ground-cancelling outputs, 98
Ground currents, 95
Ground loops, 11, 91
Grounding, 95
Group module, 128

HF stability, 159, 163

Impedance-balanced line outputs, 97
Input impedances, 69, 103, 106, 108, 122
Input selector switch, 52, 67
Instrumentation amplifiers, 111

JFETs:
  cascaded switches, 160
  changeover, 159
  control voltage feedthrough, 169

JFETs (*Continued*)
current mode, 161
series switching, 155
shunt switching, 160
Johnson noise, 54, 55, 61, 69, 110, 151

Ladder attenuator, 46
LED biasing, 1
Limiter, 117–26
Line inputs, 102
Line outputs, 97–102
Linear controls, 33, 73, 77, 135
Linearity, 40, 150
Lipshitz, Stanley, 30
Low-impedance design, 31

Magnetic coupling, 94
Microphone preamplifiers, 110, 128
Mix bus, 137
Mixdown, 141
Mixing desks, 107, 127–41
Moving-coil cartridges, 37, 54, 88
output voltages, 38, 54, 88
Moving-magnet cartridges, 38
Multiple input devices, 40
Multitrack recording, 140
Mute blocks, 162
Muting relay, 7, 9, 80

Negative feedback, 14, 43
Negative-impedance converters, 136
Noise gating, 2, 9
Noise, Johnson, 54, 55, 61, 110
Noise performance, 25, 31, 39, 47, 61–3, 70, 73, 75, 84, 140
Normalization amplifier, 6
Nyquist stability criterion, 17

Offness, 143
Opamps, 24, 34, 50, 74, 107, 150
compensation, 34, 65
Overload margin, 2, 4, 14, 55, 86–9

Panpots, 135
Passive equalisation, 57
Phantom power, 133
Phase-shift, 76
Phase-splitter, 122
Physical layout, 10, 22, 48
Post-fade sends, 135
Power supply, 10, 14, 22, 34, 82
Preamplifier:
architecture, 25, 55
distortion, 3, 14, 19
output levels, 52
Prefade-listen (PFL), 128
Prefade sends, 135
Push–Pull Class-A operation, 13

Quasi-floating outputs, 100

Recording, 140
Relay control, 80
Relays, 143
RIAA equalization, 2, 14, 19, 30, 53, 54, 57–61, 88
IEC amendment, 30, 88
Routing matrix, 141

Screened cables, 92
Series feedback in phono stage, 2
Shunt feedback in phono stage, 2
Shunt muting circuits, 150
Signal-to-noise ratio, 14
Soft changeover, 168–9
Solid-state switching, 128, 139
SPICE simulation, 74
Subsonic filter, 14, 31, 62, 140
Summing amplifiers, 137
Superbal amplifier, 108
Switched-gain balanced amplifier, 107
Switch-on transients, 161
Symmetry, 139

Tape monitoring, 69, 141
Tape outputs, 6, 50, 69
Tone cancel, 75
Tone controls, 7, 22, 26, 32, 50, 70–2, 128, 134
Transformer inputs, 104, 133
Transformer outputs, 102
Treble frequency control, 72

Unbalanced line inputs, 103
Unbalanced line outputs, 97
Unity-gain buffers, 52

Variable-gain balanced amplifier, 109
Vernier balance control, 69
Virtual-earth, 137
Virtual mixing, 141
Voltage-controlled attenuator (VCA), 118
Volume control, 3, 52, 77
  gain law, 33
  imbalance, 33
VU meters, 8

Zero-crossing detection, 80, 139, 162

## POWER AMPLIFIERS

AC coupling, 137
Amplifier classes, 404

Balance of input stage, 207
Balanced VAS, 226
Band-gap references, 352
Batman curve, 396
Bipolar beta droop, 252, 369, 371
Blameless amplifiers, 266, 319
Blomley amplifier, 406
Bootstrapping:
  collector loads, 215, 222, 393
  input, 344, 378

Capacitor distortion, 171
Cascode VAS, 174, 224
Class-A, 232, 301, 405
  quiescent current control, 349
Class-AB, 233, 406
Class-B, 233, 406
Class-C, 406
Class-D, 406
Class-G, 414, 440
Class-S, 408
Commutating amplifier, 412
Compensation, 279
  dominant pole, 195, 280
  lag, 281
  two pole, 284, 290
Complementary feedback pair (CFP), 247, 358
Conjugate curves, 182, 235, 254
Constant-$g_m$ degeneration, 212, 215
Crossover distortion, 191, 192, 232–49, 254, 324
Current-mirrors, 175, 208

Darlingtons, 245
DC offset detection, 171, 428
DC offset voltage, 175, 202, 204, 208, 347
Decoupling, 272–4
Degeneration, 206
Differential input stage, 202
Distortion, 191
Dominant pole, 189

Edwin amplifier, 410
Emitter degeneration, 206
Emitter-followers, 175

Feedback factor, 359, 383, 393, 394
Feedforward diodes, 366
FET output stage, 172, 182, 394

Gain measurement (open-loop), 196
$G_m$ doubling, 233

Harmonics, 188, 192
Hybrid BJT/FET output stages, 176

Inclusive Miller compensation, 282
Input currents, 310
Input impedance of output stage, 176, 258
Input Stage cascoding, 214
Input stages:
  bipolar, 199
  cascode, 315, 394
  Cascomp, 212, 213
  CFP input, 212, 214
  cross-quad, 212, 213
  current-mirror loading, 175, 208
  differential, 202
  FET, 314
  offset voltage, 202
  singleton, 204
Instability, high frequency, 279–84

Johnson noise, 311

Latch-up, 359
Lag compensation, 281
Large-signal non-linearity, 251–4, 328, 366
Linearity, 186, 196, 201
Load-invariant amplifier, 363
Loudspeaker currents, 400

Mains-fail detection, 426
Miller integrator, 175, 194
Model amplifiers, 198
Muting relays, 421

Negative feedback, 171, 378
Nested feedback, 282
Noise, Johnson, 311, 343
Noise performance, 342
Nyquist stability criterion, 197, 221, 359

Opamps, 154
Open-loop bandwidth, 227–8
Open-loop gain:
  calculation, 196
  measurement, 196
Oscillation, 176
Output capacitors, 204
Output stages:
  complementary feedback pair (CFP), 247
  emitter follower, 245
  FET, 248
  quasi-complementary, 247
  triples, 248

Peak power dissipation, 445–6
Peak-to-mean ratio, 448
Phase margin, 282
Phase-splitter, 395
Pole, 189, 280
Pole splitting, 281
Power amplifiers:
  blameless, 266, 287
  Blomley, 406
  Class-A, 217
Power dissipation, 434
Power-partition diagrams, 438
Power supplies, 171
Power supply rejection, 194, 274, 394
Push–pull operation, 295

Quarter-squares principle, 380
Quasi-complementary output stages:
  bipolar, 247
  FET, 176

Quiescent current:
  control in Class-A, 301
  control in Class-B, 265, 288

Reactive loads, 442
Relays, 171, 421–33
Residual, 319
Ripple rejection, 311

Singleton input stage, 204
Slew-rate, 171, 175, 215, 357, 429
Spectrum analysis, 233
SPICE simulation, 186, 188, 369, 401, 435
Stability of power amp, 279–84
Switchoff distortion, 256

Thermal compensation, 375
Thermal design, 354
Thermal distortion, 194
Total Harmonic Distortion, 319
Transconductance, 183, 201–203, 206, 211, 217, 219, 222
Transient response, 27
Transistor beta, 348
Transresistance, 189
Trimodal amplifier, 334
Turn-on transients, 424
Two pole compensation, 284
Two-stage amplifiers, 391

Underbiasing, 233

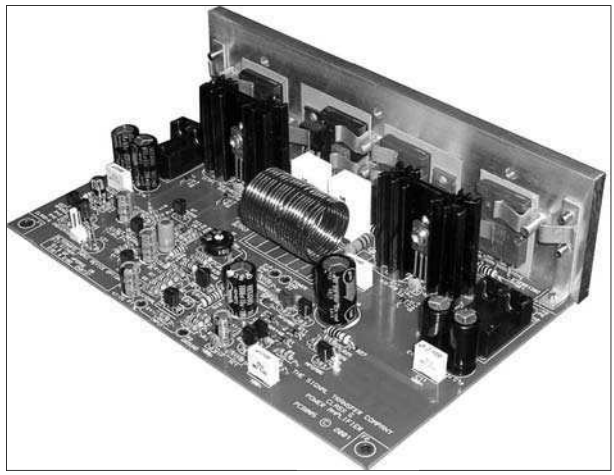
VAS buffering, 176, 225, 237
Vbe multiplier, 349
Voice-coil inductance, 401
Voltage amplifier stage (VAS), 173, 217

Warm-up time, 353

Zener clamping (of relays), 427
Zero-crossing detection, 426
Zobel network, 289

# The Signal Transfer Company

**The Signal Transfer Company is the only source for PCBs guaranteed to comply with the preamp and power amplifier design philosophies pioneered by Douglas Self**



Shown above is the Class-G power amplifier, combining improved efficiency with first-class performance. The design is described in detail in Douglas Self's *Audio Power Amplifier Design Handbook*, published by Newnes (an Imprint of Elsevier).

The following PCBs are available:

- **The Precision Preamplifier** – active gain-control and variable-frequency tone controls
- **The Load-Invariant Power Amplifier** – very low distortion into heavy loads
- **The Trimodal Power amplifier** – ultra-low distortion Class A, switchable to Class B
- **The Class-G power amplifier** – with driver circuit for Class-G indicator LED

Our PCBs have been designed with meticulous care at every point. The power amplifier board layouts are precisely the same as those approved by Douglas Self when his famous series of articles on power amplifier distortion appeared in Electronics World. You can therefore be confident that proper operation is built-in.

We supply the finest quality fibreglass PCBs, single or double-sided as appropriate. All boards have a full solder mask, tinned pads, and a silk-screen component layout. Each PCB is supplied with extensive constructional notes, previously unpublished information about the design, and a detailed parts list to make ordering components simple.

Kits of parts to build the above PCBs are also available. These contain all PCB-mounted parts, including machined heatsink coupling plates for the power amplifiers, as shown in the illustration above.

For prices and more information go to **http://www.signaltransfer.freeuk.com/** or contact:


The Signal Transfer Company
35 Hirst Grove
Dodd Naze
Hebden Bridge
West Yorkshire
HX7 8DN
United Kingdom

Tel: 01422 885196